

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_nzp model

Run on: August 28, 2004, 07:42:36 ; Search time 207.5 Seconds  
(without alignments)  
5233.801 Million cell updates/sec

Title: US-10-676-079-3  
Perfect score: 3119  
Sequence: 1 ctgaagcttcgactctccg.....atactagctcctgactcgt 1721

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame\_nzp.model -DBV=xlp  
-O=/cgn2\_1/USPTO\_spool\_p/US10676079/runat\_28082004\_084225\_23813/app\_query.fasta\_1.1863  
-DB=SPTRMBL\_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIG=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10676079@cgn2\_1.1.293@runat\_28082004\_084225\_23813 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEBUERT -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DBV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	2838	91.0	543	Q9Y251	Q9Y251 homo sapien

2	2817	90.3	545	4	Q9UL39	Q9UL39 homo sapien
3	2282	73.2	545	6	Q9MYO0	Q9MYO0 bos taurus
4	2150	68.9	535	11	Q8K3K3	Q8K3K3 mus musculus
5	2123	68.1	536	11	Q9QZP8	Q9QZP8 rattus norv
6	1645.5	52.8	523	13	Q90YK5	Q90YK5 gallus gall
7	1154.5	37.0	592	4	Q9HB37	Q9HB37 homo sapien
8	1146.5	36.8	592	4	Q8WMQ2	Q8WMQ2 homo sapien
9	1015.5	32.6	548	4	Q8WMQ1	Q8WMQ1 homo sapien
10	936.5	30.0	534	4	Q9HB38	Q9HB38 homo sapien
11	897.5	28.8	480	4	Q9HB39	Q9HB39 homo sapien
12	696	22.3	515	5	Q8T108	Q8T108 bombyx mori
13	416	13.3	521	10	Q9SDA1	Q9SDA1 arabidopsis
14	416	13.3	543	10	Q9FF10	Q9FF10 arabidopsis
15	406.5	13.0	559	16	Q89FP9	Q89FP9 bradyrhizob
16	392.5	12.6	544	10	Q8H615	Q8H615 oryza sativ
17	382	12.2	527	10	Q9LRC8	Q9LRC8 scutellaria
18	363	11.6	536	10	Q9FZP1	Q9FZP1 arabidopsis
19	352.5	11.3	516	10	Q9FLK8	Q9FLK8 arabidopsis
20	352.5	11.3	539	10	Q8L608	Q8L608 arabidopsis
21	169.5	5.4	190	10	Q82604	Q82604 arabidopsis
22	160	5.1	935	5	Q9VE79	Q9VE79 drosophila
23	135	4.3	411	16	P72895	P72895 synecocyst
24	130.5	4.2	493	17	Q9HK01	Q9HK01 thermoplasma
25	126	4.0	408	3	Q9HEZ1	Q9HEZ1 phanerochaete
26	126	4.0	408	3	Q9HEZ2	Q9HEZ2 phanerochaete
27	124	4.0	366	10	Q8GSP4	Q8GSP4 lotus japon
28	118.5	3.8	2319	3	Q96U00	Q96U00 neurospora
29	117	3.8	489	10	Q9SS90	Q9SS90 arabidopsis
30	116.5	3.7	617	12	Q40936	Q40936 measles vir
31	114.5	3.7	1169	5	Q869K5	Q869K5 dictyostell
32	114	3.7	398	16	Q8F410	Q8F410 leptospira
33	114	3.8	832	4	Q75427	Q75427 homo sapien
34	114	3.7	1449	10	Q7XPA4	Q7XPA4 oryza sativ
35	113.5	3.6	287	2	Q59144	Q59144 aeromonas b
36	113.5	3.6	617	12	Q83295	Q83295 measles vir
37	113.5	3.6	1829	2	Q9RH44	Q9RH44 pantoea agg
38	113	3.6	390	17	Q8TPH7	Q8TPH7 methanosaarc
39	113	3.6	641	17	Q97B86	Q97B86 thermoplasma
40	112.5	3.6	617	12	Q83647	Q83647 measles vir
41	112.5	3.7	692	16	Q81ZV3	Q81ZV3 streptomyces
42	112.5	3.6	1276	16	Q82J10	Q82J10 streptomyces
43	112	3.6	636	16	Q88RX6	Q88RX6 lactobacilli
44	112	3.7	680	5	Q95TU3	Q95TU3 drosophila
45	112	3.7	1224	5	Q9VXW2	Q9VXW2 drosophila

#### ALIGNMENTS

RESULT 1  
ID Q9Y251 PRELIMINARY; PRT; 543 AA.

AC Q9Y251, 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE HEPARANASE (HPSB protein).  
GN HPA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=95321249; PubMed=10395326;  
RA Hulst M.D., Freeman C., Hamdorf B.J., Baker R.T., Harrie M.J., Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis";  
RL Nat. Med. 5:803-809(1999).  
RN (2)  
RP SEQUENCE FROM N.A.  
RA Vlodevsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,

RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
RA Spector L., Becker I.;  
RT "Mamalian heparanase: a novel gene involved in tumor progression and  
RT metastasis";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99377052; PubMed=10446189;  
RA Toyoshima M., Nakajima M.;  
RT "Human heparanase. Purification, characterization, cloning, and  
RT expression";  
RL J. Biol. Chem. 274:24153-24160(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=9935379; PubMed=10405343;  
RA Kusie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,  
RA Seddon A.P., Giorgio N.A., Bohlen P.,  
RT "Cloning and Functional Expression of a Human Heparanase Gene";  
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datchenko L., Marziani K., Farmer A.A., Rubin G.M., Hong L.,  
RA Brownstein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliyil S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzyzanski M.I., Skolnik U., Smallegre D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF165154; AAD45379.1; -  
DR EMBL: AF144325; AAD41342.1; -  
DR EMBL: AF155510; AAD54941.1; -  
DR EMBL: AF152376; AAD54669.1; -  
DR EMBL: BC051321; AAH51321.1; -  
DR Genev. HGNC:5164; HPSR  
DR GO: GO:0004566; F:beta-glucuronidase activity; TAS.  
DR GO: GO:0007125; P:invasive growth; TAS.  
DR GO: GO:0006029; P:proteoglycan metabolism; TAS.  
DR InterPro: IPR005199; Glyco hydro 79N.  
DR Pfam: PF03662; Glyco hydro 79n; I.  
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

## Alignment Scores:

Pred. No.: 6.3e-222 Length: 543  
Score: 2838.00 Matches: 542  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.82% Mismatches: 0  
Query Match: 90.99% Indels: 0  
DB: 4 Gaps: 0

US-10-676-079-3 (1-1721) X Q9Y251 (1-543)

QY 63 ATGCTGCTGCGCTGGAAGCTGCGCTGCCGCCGCGCTGATGCTGCTCTGCGGCGG 122

Db |||||  
QY 1 MetLeuLeuArgSerIysProAlaLeuProProLeuMetLeuLeuLeuGlyPro 20  
123 CTGGATCCCTCCCTCCCTGGCGCCCTGCGCCGACCTGCGCAAGACAGGACCGTGGAGC 182  
Db |||||  
QY 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnValValAsp 40  
183 CTGGACTTCTTCCACCGAGAGCGCTGCACTGGTAGGCCCTCGTCTGTCCTCCGACCC 242  
Db |||||  
QY 41 LeuAspPhePheThrGlnGlnProLeuHLeuValSerProSerPheLeuSerValThr 60  
QY 243 ATTGAAGCCCAACCTGGCCACGAGACCCGCGGTTCTCTATCTCTGGGTTCTCCAAAGCTT 302  
Db |||||  
QY 61 LeAspAlaAsnLeuAlaThrAspProArgPheLeuHLeuLeuGlySerProLysLeu 80  
303 CGTACCTTGGCCGAGAGCTTGTCTCTGCGTACCTAGGATTGGTGGACCAAGACAGAC 362  
Db |||||  
QY 81 ArgThrLeuAlaArgGlyLeuSerProAlaIylLeuAlaArgPheGlyGlyThrLysThrAsp 100  
QY 363 TTCTTAATTTTCGATCCCAAGAGATCAACCTTGAAGAGAAATTACTGGCAATCT 422  
Db |||||  
QY 101 PheLeuHLeuPheAspProLysGlySerThrPheGlyGlyLysSerThrTrpGlnSer 120  
423 CAAGTCAACCAAGATTTTGGCAAAATATGATCATCTCCCTCGATGGAGAGAGAGATTA 482  
Db |||||  
QY 121 GlnValAsnGlnAspIleCysLysThrGlySerIleProProAspValGlnGlyLysLeu 140  
483 CGGTTGAATGGCGCCCTACCAAGAGCAATGCTACTCCGAGAAACATCAACAGAAAGTTTC 542  
Db |||||  
QY 141 ArgLeuGlnIleTrpProIleGlnGlnLeuLeuAlaGlnHleThrGlnLysLysPhe 160  
543 AAGAACAGCACTACTCAAGAAAGCTGTGATGTGCTATACACTTTTGCAAATGCTCTCA 602  
Db |||||  
QY 161 LysAsnSerThrLysSerIysSerValAspValLeuThrPheAlaAsnGlySer 180  
603 GAGCTGACTGATCTTTGGCCCTTAATGGCTTATTAAGAACGACAAATTGCACTGAGAAC 662  
Db |||||  
QY 181 GlyLeuAspLeuIlePheGlyLeuAlaLeuLeuAlaLeuLysThrAlaAspLeuGlnTrpAsn 200  
663 AGTTCTAATGCTCAGTTGCTCTGAGCTGACTGCTCTCCAAAGGGATTAACATTTCTGG 722  
Db |||||  
QY 201 SerSerAsnAlaGlnLeuLeuLeuAspIleCysSerSerLysGlyThrAsnIleSerTrp 220  
723 GAACTAGGCAATGAACCTTAACAGTTTCTTAAGAAAGCTGTATTTTCATCAATGGGTCG 782  
Db |||||  
QY 221 GlnLeuGlyAsnGlnProAsnSerPheLeuLysValAlaAspIlePheHleAsnGlySer 240  
783 CAGTTAGGAGAAATATATATTCATTTGCAATTAACCTTCAAGAAAGCTTCAAAAT 842  
Db |||||  
QY 241 GlnLeuGlyGlnAspPheHleGlnLeuHleLysLeuLeuArgLysSerThrPheLysAsn 260  
843 GCAAAACCTATGCTGCTATGTTGCTGAGCTGAGCTGCAAGAAAGCGCTAAGATGCTGAAG 902  
Db |||||  
QY 261 AlaLysLeuIleCylProAspValGlyGlnProArgArgLysThrAlaLysHleLeuLys 280  
903 AGCTTCTTAAGAGCTGTGAGAGAGATGATTGATTCACTTACATGATCATCACTATTTTG 962  
Db |||||  
QY 281 SerPheLeuLysAlaGlyGlyValIleAspSerValThrPheHleIleThrLysLeu 300  
963 AATGAGCGACGCTGCTACCGAGGAAGATTTCTTAACCGATGATTAATTTGAATTTTAT 1022  
Db |||||  
QY 301 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320  
1023 TCATGCTGCAAAAAGTTTCCAGGTGTGTGAGAGACAGGCTGGCAAGAAAGCTCTGG 1082  
Db |||||  
QY 321 SerSerValGlnLysAlaPheGlnValIleGlnSerThrArgProGlyLysValTrp 340  
1083 TTAGAGAAACAAGCTCTGCATATGAGAGCGGAGCGCTTGTCTATCCAGACCTTTGCA 1142  
Db |||||  
QY 341 LeuGlyGlyThrSerSerAlaThrGlyGlyAlaProLeuLeuSerAspThrPheAla 360  
1143 GCTGGCTTATATGAGCTGATTAATTTGGCTGTCAAGCCGCAATGGGAATGAAGTGTG 1202

```

Db      361 AlaglypHemeEtripleuaspHysLeuGlyLeuSerAlaArgMetGlyIleGluVal 380
QY      1203 ATAGGCAAGTATTCCTTGGACAGAACTACATTTAGTGATGAAGAACTTCATCT 1262
Db      381 MetAlaGlnValPhePheGlyAlaGlyAsnThrHisLeuValAspGlnAsnPheAspPro 400
QY      1263 TTACCTGATTATTTGGCTATCTCTTCTGTTCAAGAAATTTGGTGGCCCAAGGTGTTAATG 1322
Db      401 LeuProAspTyrTrpLeuSerLeuLeuPheLysLeuValGlyThrIleValLeuMet 420
QY      1332 GCAAGCGTGCAGGTTCCAAAGAGAGAAAGCTTCAGATTTACCTTATTCGCAAAACACT 1382
Db      421 AlaSerValGlnGlySerIleAspArgLysLeuArgValTyrLeuHisCysThrAsnThr 440
QY      1383 GACATTCACAGGTAATAAGAGAGATTAACCTGTATGCAATAAAGCTTCATACGTC 1442
Db      441 AspAsnProArgTyrIleGlySerGlnGlyAspLeuThrLeuTyrAlaIleAsnLeuHisAsnVal 460
QY      1443 ACCAAGTACTTCGCGTTACCTTACCTTCTTAAACAGCAAGTGATTAATACCTTCTTA 1502
Db      461 ThrIleTyrLeuArgLeuProTyrProPheSerAsnLysGlnValAspLysTyrLeuLeu 480
QY      1503 AGACCTTTGGACCTTCATGATTTACTTCCAAATCTGTCCAACTCAATGCTTAACTTA 1562
Db      481 ArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeuThrLeu 500
QY      1563 AAGATGTGTGATGATCAACCTTGCACCTTTAATGGAAGAAAGCTTCGCGGCGAGAGAGT 1622
Db      501 LysMetValAspAspGlnThrLeuProPheLeuMetGlnLysProLeuArgProGlySer 520
QY      1623 TCACCTGGGCTTCGACGCTTTCATATAGTTTGTGTATGAAGATCCAAAGTTGGCT 1682
Db      521 SerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLysValAla 540
QY      1683 GCTTGATC 1691
Db      541 AlaCysIle 543

```

## RESULT 2

```

Q9ULJ39 PRELIMINARY; PRT; 545 AA.
ID      Q9ULJ39
AC      Q9ULJ39;
DT      01-MAY-2000 (TRMBLrel. 13, Created)
DT      01-MAY-2000 (TRMBLrel. 13, Last sequence update)
DT      01-JUN-2002 (TRMBLrel. 21, Last annotation update)
DE      Heparanase.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Placenta;
RX      MEDLINE=20229546; PubMed=10764835;
RA      Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RT      "Heparanase expression in invasive trophoblasts and acute vascular
RT      damage.";
RL      Glycobiology 10:467-475(2000).
DR      EMBL; AF084467; AAD54516.1; -.
DR      InterPro; IPR005199; Glyco_hydro_79N.
DR      Pfam; PF03662; Glyco_hydro_79n; 1.
SQ      SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;

```

## Alignment Scores:

```

Pred. No.: 3,21e-220 Length: 545
Score: 2817.00 Matches: 542
Percent Similarity: 99.63% Conservative: 1
Best Local Similarity: 99.45% Mismatches: 0
Query Match: 90.32% Indels: 2
DB: 4 Gaps: 2

```

US-10-676-079-3 (1-1721) x Q9ULJ39 (1-545)

```

QY      63 ATCTGCTGCGCTCGAAGCTGCGCTGCGCGCGCG---CTGATCTGCTGCTCTGGG 119
Db      1 MetLeuLeuArgSerLysPheProAlaLeuProProProLeuMetLeuLeuLeuGly 20
QY      120 CCGCTGGGTCCCTCTCTCCCTGGCGCCCTGCCCCGACCTGCGCAAGCA---CAGAGCTC 176
Db      21 ProLeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnGlnAspVal 40
QY      177 GTGAGCTCGGACTTCTTCAACCCAGAGACCGGCGACCTGGTAGGCCCCGCTCCGTGCC 236
Db      41 ValAspLeuAspPhePheThrGlnGlnProLeuHisLeuValSerProSerPheLeuSer 60
QY      237 GTCAACATTGACGCCAACCTGCGCACAGACCCGCGTTCTTCATCTCTGCGTTCTCCA 296
Db      61 ValThrIleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerPro 80
QY      297 AAGCTTCGTAACCTTGACCAAGAGGCTTGTCTCTGCGTACCTGAGGTTGGTGGACCAAG 356
Db      81 LysLeuArgThrLeuAlaArgGlyLeuSerProAlaTyrLeuArgPheGlyGlyThrLys 100
QY      357 ACAGACTTCCTAATTTGATGCCAAGAAAGAAATCACTTGAAGAGAGAGATTAATCG 416
Db      101 ThrAspPheLeuIlePheAspProLysGlnSerThrPheGlnGlnArgSerTyrTrp 120
QY      417 CAATCTCAAGTCAACCAAGATATTTGCAAAATATGATCCATCCCTCTGATGTGAGAGAG 476
Db      121 GlnSerGlnValAsnGlnAspIleCysLysTyrGlySerIleProProAspValGlnGln 140
QY      477 AAGTTACGGTTGGAATGGCCCTTACCAAGAGCAATTGCTACTCGAGAAACATCAACAGAA 536
Db      141 LysLeuArgLeuGlnTyrProTyrGlnGlnGlnLeuLeuArgGlyHisTyrGlnLys 160
QY      537 AAGTTCAGAAAGAGACCTTACCAAGAGCTGTGATGATGCTATPACCTTTGCAAG 596
Db      161 LysPheLysAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsn 180
QY      597 TECTCAGACTGAGCTTGATCTTTGGCCTTAATGGCTTATTAAGAACAGCAGATTGTCAG 656
Db      181 CysSerGlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGln 200
QY      657 TCGAAGAGTTTAATGCTCAGTTGCTCTGCACTACCTGCTTCCAGGGGATTAACATT 716
Db      201 TrpAsnSerSerAsnAlaGlnLeuLeuAspTyrCysSerSerLysGlyTyrAsnIle 220
QY      717 TCTTGGAAGCTGAGGAAGAACTTAAGCTTCCCTTAAGAGCGATATTTTCATCAT 776
Db      221 SerTyrGlnLeuGlyAsnGlnProAsnSerPheLeuLysValaAspIlePheIleAsn 240
QY      777 GGGTCGAGTTAGAGAGATTAATATTCATTAATGCAATTAACCTTAAGAAAGTCCACCTTC 836
Db      241 GlySerGlnLeuGlyGlnAspPheIleGlnLeuHisLysLeuLeuArgLysSerThrPhe 260
QY      837 AAAAAATGCAAACTGTATGCTCTGATGTTGGTCAGCTTCGAAGAAGACGGCTAAGATG 896
Db      261 LysAsnAlaLysLeuTyrGlyProAspValGlyGlnProArgArgLysThrAlaLysMet 280
QY      897 CTGAAGAGCTTCTGAAGGCTGTGAGAGAGATGATGATTCAGTTACATGCAATCACTAC 956
Db      281 LeuLysSerPheLeuLysAlaGlyGlyValIleAspSerValThrTrpHisTyr 300
QY      957 TATTTGAATGACAGCACTGCTCCAGGGAAGATTTCTAAACCTGATGATATGACATT 1016
Db      301 TyrLeuAsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIle 320
QY      1017 TTTATTTCACTGTGCAAAAAGTTTTCAGGTGTGAGAGACCAAGGCTGCGCAAGAG 1076
Db      321 PheIleSerSerValGlnLysValPheGlnValaGlnLysThrArgProGlyLysLys 340
QY      1077 GTCGTGTTAAGAGAAACAAGCTTCGATATGAGAGCGGAGCGCCCTTGATCCAGACCC 1136
Db      341 ValTrpLeuGlyGlnThrSerSerAlaTyrGlyGlyAlaProLeuLeuSerAspThr 360
QY      1137 TTTGACAGCTGAGCTTATATGCTGATTAATTTGGCTGACGCCGATGGAATATAGA 1196

```

```

Db      361 PheAlaIaGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaArgMetGlyIleGlu 380
QY      1197 GTGTGATGAGGCAAGATATCTTTGGAGCGAGAAACCTAGCATTTAGTGATGAGAAACTTC 1256
Db      381 ValValMetArgGlnValPhePheGlyAlaGlyAsnThrIleuValAspGluAspPhe 400
QY      1257 GATCCTTACCTGATTAATTTGGCTATCTCTCTGTTCAAGAAATTTGGTGGCCACCAAGTG 1316
Db      401 AspProLeuProAspTyrTrpLeuSerLeuLeuPheLysLeuValGlyThrVal 420
QY      1317 TTTAATGGCAGCGGTGCAAGGTTCAAGAGAAAGAGCTTGAGTATACCTTCATTGACA 1376
Db      421 LeuMetLaserValGlnGlySerLysArgArgLysLeuArgValTyrLeuH1sCySerThr 440
QY      1377 AACACTGACATCCAGGATTAATAAGAGAGATTAATCTGTATGCGCATTAACCTCAT 1436
Db      441 AsnThrAspAsnProAlaGlyTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLeuH1s 460
QY      1437 AACGTCACCAAGTACTTGGGTTACCTCATCTTTTCTAACAAGAGAGTGAATAATAC 1496
Db      461 AsnValThrLysTyrLeuAlaGlyLeuProTyrProPheSerAsnLysGlnValAspLysTyr 480
QY      1497 CTTCTAAGACCTTTGGGACCTCATGATTAATCTTTCCAAATCTGTCACACTCATGCTCTA 1556
Db      481 LeuLeuArgProLeuGlyProH1sGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeu 500
QY      1557 ACTCTAAGAGTGGTGAATGATGATCAAAACCTTGCGACCTTTAATGAGAAACCTCCGGACA 1616
Db      501 ThrLeuLysMetValAspAspGlnThrLeuProLeuMetGlnLysProLeuAspPro 520
QY      1617 GGAAGTTCAGTGGGCTTGCCAGCTTCTCATATATGTTTTTGTGATAGAAATGCCAAA 1676
Db      521 GlySerSerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLys 540
QY      1677 GTTGCTGCTTGCAATC 1691
Db      541 ValAlaIaIaCysIle 545

RESULT 3
Q9MYT0 ID Q9MYT0 PRELIMINARY; PRT; 545 AA.
AC Q9MYT0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
OS Heparanase.
DS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of Heparanase mRNA in Bovine Placenta During Gestation.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF81160; AAF87301.2; -
DR InterPro: IPR005199; Glyco_hydro_79N.
DR Pfam: PF03662; Glyco_hydro_79n; I.
SQ SEQUENCE 545 AA; 61076 MW; FAQ4BDFPD85B933 CRC64;

```

Alignment Scores:

Pred. No.:	9_98e-177	Length:	545
Score:	2282.00	Matches:	435
Percent Similarity:	86.24%	Conservative:	35
Best Local Similarity:	79.82%	Mismatches:	73
Query Match:	73.16%	Indels:	2
DB:	6	Gaps:	1

US-10-676-079-3 (1-1721) x Q9MYT0 (1-545)

```

QY      63 ATGCTGCTGGCTCGAAGCCTGCGTCCGCGCGCTGATGCTGCTG-----CTCTTG 116
Db      1 MetLeuAlaCysArgLysProGlyLeuArgProProLeuLeuLeuLeuProLeuLeu 20
QY      117 GGGCGCGCTGGTCCCTCTCCCTGGCGCGCTCCCGGACCTGGCGGACACAGCGTCC 176
Db      21 GlyProLeuLysProCysSerProGlyThrProAlaAlaAlaAlaProAlaAspAla 40
QY      177 GTGAGCTGGACTTTCTTCAACCCAGAGCCGCTGACCTGGTGAAGCCCTGTTCTGTC 236
Db      41 AlaGluLeuGlnPhePheThrGluArgProLeuH1sLeuValSerProAlaPheLeuSer 60
QY      237 GTCAACATTAAGCGCAACCTGGCCACGAGACCGCGGTTCTCATCTCTCGGTTCCA 296
Db      61 PheThrIleAspAlaAsnLeuAlaThrAspProArgPhePheThrPheLeuGlySerSer 80
QY      297 AAGCTTCGATCCTTGGCGCAGAGCTGTGTCCTCCGCTGACCTGAGCTTGGTGGCCACCA 356
Db      81 LysLeuArgThrLeuAlaArgGlyLeuAlaProAlaTyrLeuArgPheGlyGlyAsnLys 100
QY      357 ACAGACTTCCTAATTTTCGATCCCAAGAGAAATCAACCTTTGAGAGAGAAATTAAGT 416
Db      101 GlyAspPheLeuIlePheAspProLysGlyGlnProAlaPheGluGlnArgSerTyrTrp 120
QY      417 CAATCTCAAGTCAACCAAGATATTGCAAAATATGATTCATCCCTCTGATGTGAGAG 476
Db      121 LeuSerGlnSerAsnGlnAspIleCysLysSerGlySerIleProSerAspValGlnGlu 140
QY      477 AAGTTACGGTTGGAAGGCGCTTACCGAGAGCAATGCTCTCCGAGAAACCTACAGAA 536
Db      141 LysLeuArgLeuGlnTrpProPheGlnGlnValLeuLeuArgGlnGlnTyrGlnLys 160
QY      537 AAGTTCAAGAACAGCACTACTCAAGAAAGCTGTGATGTGCTATATCACTTTGCAAC 596
Db      161 LysPheThrAsnSerThrTyrSerArgSerSerValAspMetLeuThrPheAlaSer 180
QY      597 TGCTCAAGACTGACTTGAATCTTGGCCTTAATGCTTATTAAGAACAGCATTTGCA 656
Db      181 CysSerGlyLeuAsnLeuIlePheGlyValAsnAlaLeuLeuArgThrThrPheMetH1s 200
QY      657 TGGACAGTTCATATGCTCAGTGGCTCCGAGACTAGTGCCTTCCAAAGGGGTAAACAT 716
Db      201 TrpAspSerSerAsnAlaGlnLeuLeuAspTyrCysSerSerLysAsnTyrAsnIle 220
QY      717 TCTTGGAACTAAGCAATGAACCTTAACGTTTCTTAAGAAAGCTGATATTTTCATCA 776
Db      221 SerTrpGluLeuGlyAsnGlnProAsnSerPheGlnArgLysAlaGlyIlePheIleAsn 240
QY      777 GGGTCCGAGTTAGAGAAAGATTAATTAATTCATTAAGCAATTTTAAGAAAGTCCACT 836
Db      241 GlyArgGlnLeuGlyGlnAspPheIleGlnPheArgLysLeuLeuGlyLysSerAlaPhe 260
QY      837 AAAAATGCAAAACTCATGCTGCTGATGTTGGTGCAGCTCGAAGAAAGCGGTAAAG 896
Db      261 LysAsnAlaLysLeuTyrGlyProAspIleGlyGlnProArgArgAsnThrValLysMet 280
QY      897 CTGAAGAGCTTCTGAAGCGCTGTGAGAAAGTATGATTAAGTTCATGTCATCACTAC 956
Db      281 LeuLysSerPheLeuLysAlaGlyGlyValIleAspSerValThrTrpHisnIleTyr 300
QY      957 TATTGAATGACGAGCTGCTACACGAGAAAGATTTTCTAAACCTGATGATTTGACAT 1016
Db      301 TyrValAsnGlyArgIleAlaThrLysGlnAspPheLeuAsnProAspIleLeuAspThr 320
QY      1017 TTTAATTCATCTGTGAAGAAAGTTTCCAGTGCGTGAAGACACAGGCGCTGGCAAGAG 1076
Db      321 PheIleSerSerValGlnLysThrLeuArgIleValGlnLysIleArgProLeuLysLys 340
QY      1077 GTCTGTTAGAGAAACAGCTCTGATATGAGAGCGAGCGGCTTGTGATTCGACAC 1136
Db      341 ValTrpLeuGlyGlnThrSerSerAlaPheGlyGlyAlaProPheLeuSerAsnThr 360
QY      1137 TTTCAGCTGGCTTTATGTGGCTGATTAATTTGGGCGCTGTCAAGCCGCAATGGAA 1196

```

```

Db 361 PheAlaAlaGlyPheMetTrpLeuSerPheGlyLeuSerAlaArgMetGlyLeu 380
Qy 1197 GGGGATGATGAGCAAGATATCTTTGGAGACAGAAATCACTTATGATGATGAAAACCTTC 1256
Db 381 ValValMetArgGlyValLeuPheGlyAlaGlySerTrpHisLeuValAspGlyAsnPhe 400
Qy 1257 GATCCCTTACCTGATATATGCGATCTCTTCTGTTCAAGAAATGGTGGCCCAAGGCG 1316
Db 401 GTPProLeuProAspTrpTrpLeuSerLeuPheLeuValAlaGlyAsnVal 420
Qy 1317 TTAATGGCAAGCGGTGCAAGTTCAAGAGAAAGCTTCAAGTATACCTTCAATGGCACA 1376
Db 421 LeuMetAlaSerValIleGlyProAspArgSerLysPheArgValTrpLeuHisCysThr 440
Qy 1377 AACACTGACAAATCCAAAGTTATAAGAGAGATTTACTCTGATGCCATAAACCTCCAT 1436
Db 441 AantTrpLysHisProArgTrpLysGlyAspLeuThrLeuLysTrpHisLeuHis 460
Qy 1437 AACGTCACCAAGATCTTGGGTTACCTTACCTTTTCTTCAACAGCAAGTGATAAATAC 1496
Db 461 AantValThrLysHisLeuGlyLeuProHisHisLeuPheAsnLysGlyValAspLysTrp 480
Qy 1497 CTTTCAAGACTTTGGGACCTCAATGATTAATCTTCCAAATCTGTCCAACTCAATGCTCTA 1556
Db 481 LeuIleLysProSerGlyThrAspGlyLeuLeuSerLysSerValGlnLeuAsnGlyGln 500
Qy 1557 ACTCTTAAGATGATGATGATCAAACTTGCCCACTTTAATGAAAACCTCTCGGCGCA 1616
Db 501 IleLeuLysMetValAspGlyGlnThrLeuProAlaLeuThrGlyLysProLeuHisPro 520
Qy 1617 GGAAGTCTACCTGGGCTTGGCAGCTTCTCATATAGTTTGTGATAGAAATGCCAA 1676
Db 521 GlySerSerLeuGlyMetProProPheSerTrpGlyPhePheValIleArgAsnAlaLys 540
Qy 1677 GTTGCTGCTTGCAATC 1691
Db 541 ValAlaAlaCysIle 545

RESULT 4
ID Q8K3K3 PRELIMINARY; PRT; 535 AA.
AC Q8K3K3;
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hepatanae.
GN HPSR OR HPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=FVB;
RC MEDLINE=2354683; PubMed=12466851;
RA The PANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK047467; BAC30600.1; -
DR MGD; MGI:1343124; Hmpe.
DR InterPro; IPR005199; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79n; I.

```

```

SQ SEQUENCE 535 AA; 60065 MW; 6E73A8302FB8A0DF CRC64;
Alignment Scores:
Pred. No.: 5,336-166 Length: 535
Score: 2150.00 Matches: 407
Percent Similarity: 86.25% Conservative: 51
Best Local Similarity: 76.65% Mismatches: 73
Query Match: 68.93% Indels: 0
DB: 11 Gaps: 0

US-10-676-079-3 (1-1721) x Q8K3K3 (1-535)
Qy 99 CTGATGCTGCTCTCTGCGGCGCTGAGTCCCTCTCCCGTGGCCGCTCCGACCT 158
Db 5 LeuLeuLeuTrpLeuTrpGlyProLeuGlyAlaLeuAlaGlnGlyAlaProAlaGlyThr 24
Qy 159 GCGCAAGCAACAGACGTGTGGACCTTGACCTTTCACCCAGAGACCGCTGACCTGGTG 218
Db 25 AlaProThrAspAspValValAspLeuGlyPheTrpLysArgProLeuArgSerVal 44
Qy 219 AGCCCTGCTTCTGCTCCGTCACATTAAGCGCAAGCTGGCCAGACCGCGGCTTCTC 278
Db 45 SerProSerPheLeuSerIleThrIleAspAlaSerLeuAlaThrAspProArgPheLeu 64
Qy 279 ATCTCTGCGGTTCTTCAAGCTTGTACTCTTGCCAGAGGCTTGTCTCTGCTGACTG 338
Db 65 ThrPheLeuGlySerProArgLeuArgAlaLeuAlaArgGlyLeuSerProAlaTrpLeu 84
Qy 339 AGGTTTGGTGGCAACAGACACTTCTTAATTTTGATCCCAAGAGAAATCAACTTT 398
Db 85 ArgPheGlyGlyThrLysThrAspPheLeuIlePheAspProAspLysGlyProThrSer 104
Qy 399 GAAGAGAGAGTACTGCGAATCTTCAAGTCCCAAGCAAGATTTGGAATATGATCCATC 458
Db 105 GlnGluArgSerTrpLysSerGlnValAsnHisAspIleCysArgSerGluProVal 124
Qy 459 CCTCTGATGTGAGAGAGATTAAGCTTGAATGAGCCCTACAGAGACATTTGCTACTC 518
Db 125 SerAlaAlaValLeuArgLysLeuGlnValGluTrpProPheGlnGluLeuLeuLeu 144
Qy 519 CGAAGAACTACACAGAAAAAGTTCAAGAACACAGACTTACTCAAGAACTCTGATAGTGTG 578
Db 145 ArgGlnGlnTrpGlnLysGlnLysLeuPheLysAsnSerThrTrpSerArgSerSerValAspMet 164
Qy 579 CTAATCACTTTGCAAACTGCTCAGACGTGCACTTGAATCTTGGCTAAATGCTTAATA 638
Db 165 LeuTrpSerPheAlaLysCysSerGlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeu 184
Qy 639 AGAAGACAGATTTGACGTGAGAACAGTTCTAATGCTCAATGCTCTGCTGACTACTGCT 698
Db 185 ArgThrProAspLeuArgTrpAsnSerSerAlaGlnLeuLeuLeuAspTrpCysSer 204
Qy 699 TCCAAAGGGGATTAACATTTCTTGGAACTAGGCAATGAACCTTAACAGTTTCTTAAGAG 758
Db 205 SerLysGlyTrpAsnIleSerTrpGluLeuGlyAsnGluProAsnSerPheTrpLys 224
Qy 759 GGTGATATTTTATGAAAGCGTGCAGATTAAGAGAAATTAATTCATTTGATTAACCT 818
Db 225 AlaHisIleLeuIleAspLysLeuGlnLeuGlyGluAspPheValGluLeuHisLysLeu 244
Qy 819 CTAAGAAAGTCCACCTTCAAAATGCAAACTCTAATGCTCAATGCTGATTTGCTGACCTCGA 878
Db 245 LeuGlnArgSerAlaPheGlnAsnAlaLysLeuTrpGlyProAspIleGlyGlnProArg 264
Qy 879 AGAAAGCGGCTTAAGATGCTGAAGGCTTCTGAAGGCTGTGAGAAAGTGAATTAATCA 938
Db 265 GlyLysThrValIleLeuLeuArgSerPheLeuLysAlaGlyGlyValIleAspSer 284
Qy 939 GTTAATGCGCACTTACATTTGAATGAAGCAGACTGTCAACAGAGAAATTTTAAAC 998
Db 285 LeuThrTrpHisIleTrpLysLeuAsnGlyArgIleAlaThrLysGlnAspPheLeuSer 304
Qy 999 CCGATGATGATGACATTTTATTTCAATCTGTGCAAAAAAGTTTTCAGAGTGGTTGAGAGC 1058

```

```

Db      305 SerAspValLeuSprThrPheIleLeuSerValGlnIlySleuLysValThrLysGlu 324
Qy      1059 ACCAGGCGCTGGCAAGAGGTCTGGTTAGAGAAACAAGCTCTGCAATGGAGCGCGAGCG 1118
Db      325 IleThrProGlyLysLysValThrPheGlyGlnIlySerSerAlaIyrgLysGlyAla 344
Qy      1119 CCCTTGCTATCCGACACCTTGGACGTGGCTTTATGTGGCTGTGATTAATTTGGGCTGTCA 1178
Db      345 ProLeuLeuSerAenThrPheAlaIagLysPheMetTrpLeuAepLysLeuGlyLeuSer 364
Qy      1179 GCGCGAATGGGAATAGAAAGTGTGATAGAGCAATATTTCTTTGAGAGCAAGAACTTACAT 1238
Db      365 AlagIlnMetGlyIleGlyValAlaMetArgGlnValPhePheGlyAlaGlyAsnTyrHis 384
Qy      1239 TTATGGATGAAAAAATTGATTCCTTACCTGATTAATTTGGCTATCTCTTCTGTTCAAGAAA 1298
Db      385 LeuValAlaPheGlnAenPheGluProLeuProAepLysTrpLeuSerLeuLeuPheLys 404
Qy      1299 TTGGTGGGCAACAAGTGTAAATGGCAAGCGTGCAGGTTCAAGAGAAAGCTTGCAG 1358
Db      405 LeuValGlyProArgValLeuLeuSerArgValIlyGlyProAepArgSerLysLeuArg 424
Qy      1359 GTATTAATCTTATGCAACAACACTGCAATCCAGATGATTAAGAGAGAGATTTAACTGTG 1418
Db      425 ValTyrLeuHisCysThrAenValTyrHisPheProArgTyrGlnGlnGlyAepLeuThrLeu 444
Qy      1419 TATGGCAATAAACCCTCAATACGTGCACCAAGTACTGGGCTTACCCCTTCTTTTCTTAAC 1478
Db      445 TyrValLeuAenIleuHisAenValThrLysHisLeuLysValProProProLeuPheArg 464
Qy      1479 AAGCAATGTGATTAATACCTTTTGAACCTTTGGACCTGATGATTACTTTCCAAATCT 1538
Db      465 LysProValAepThrTyrLeuLeuLysPheSerGlyProAepGlyLeuLeuSerLysSer 484
Qy      1539 GTCCAAATCAATGCTCTTAATCTTAAGACCTTTGGACCTGATGATTACTTTCCAAATCT 1598
Db      485 ValGlnLeuAenGlnGlnIleLeuLysMetValAepLysGlnThrLeuProAlaLeuThr 504
Qy      1599 GAAAAACCTCTCCGCGCAGCAAGTCACTGGGCTTTGGACCTTTCTCATATAGTTTCTT 1658
Db      505 GlnLysProLeuProAlaGlySerAlaLeuSerLeuProAlaPheSerLysGlyPhePhe 524
Qy      1659 GTGATTAAGAAATGCCAAAGTGTGCTGTTGCATC 1691
Db      525 ValIleArgAenAlaLysIleAlaLysIle 535

RESULT 5
Q9QZF8      PRELIMINARY;      PRT;      536 AA.
ID      Q9QZF8;
AC      Q9QZF8;
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Heparanase.
GN      HEP.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Podiyana K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT      "Heparanase from parathyroid cell line."
RL      Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF184967; AAF04563.1; -
DR      InterPro: IPR005199; Glyco_hydro_79N.
DR      Pfam: PF03662; Glyco_hydro_79n; I.
SQ      SEQUENCE 536 AA; 60569 MW; 6208B1PD9E28421 CRC64;

Alignment Scores:
Pred. No.:      8 36e-164      Length:      536
Score:      2123.00      Matches:      405

```

```

Percent Similarity: 85.23%
Best Local Similarity: 75.07%
Query Match: 68.07%
DB: 11
Gaps: 0

US-10-676-079-3 (1-1721) x Q9QZF8 (1-536)

Qy      87 CTCGCGCGCGCGGTGATGCTGTGCTCTCGGGGCGGTGGCTCCCTCCCTGGCGCGC 146
Db      2 LeuArgProLeuLeuLeuLeuTrpLysArgLysValAlaLeuThrGlnGlyThr 21
Qy      147 CTCGCGCGCGCGGTGATGCTGTGCTCTCGGGGCGGTGGCTCCCTCCCTGGCGCGC 206
Db      22 ProAlaGlyThrAlaProThrLysAepValAlaAepLeuGlnPheThrLysArgLeu 41
Qy      207 CTCGCGCGCGCGCGGTGATGCTGTGCTCTCGGGGCGGTGGCTCCCTCCCTGGCGCGC 266
Db      42 PheGlnSerValSerProSerPheLeuSerIleThrIleAepAlaSerLeuAlaThrAep 61
Qy      267 CCGCGGTTCTCATCCTCTCGGGTTCGCAAGCTTCGATCCTTGCGCAGAGGCTTGTCT 326
Db      62 ProArgPheLeuThrPheLeuGlnGlySerProArgLeuArgAlaLeuAlaArgLysLeuSer 81
Qy      327 CCGCGGTTCTCATCCTCTCGGGTTCGCAAGCTTCGATCCTTGCGCAGAGGCTTGTCT 386
Db      82 ProAlaTyrLeuArgPheGlnGlyThrLysTrpAepPheLeuIlePheAepProAenLys 101
Qy      387 GAATCAACCTTTGAAGAGAGAGTTCGCAATCTCAAGTCAACGACAGATATTTCGAAA 446
Db      102 GlnProThrSerGlnGlnLysSerLysTrpGlnSerGlnAepAenAepIleCysGly 121
Qy      447 TATGATCCATCCCTCTGATGTGAGAGAGTTCGCAAGCTTCGATCCTTGCGCAGAG 506
Db      122 SerGlnArgValSerAlaAepValLeuArgLysLeuGlnMetGlnTrpProPheGlnGln 141
Qy      507 CAATTGCTACTCCGAAACAATPACGAAAGAGTTCGCAATCTCAAGTCAACGACAGATATTTCGAAA 566
Db      142 LeuLeuLeuLeuArgGlnGlnIlyGlnArgGlnPheLysAenSerThrLysSerAysSer 161
Qy      567 TCTGTAGATGTGTATATACCTTTTGAACCTGCTCAGACTGGAATTTGATCTTTGGGCTTA 626
Db      162 SerValAepMetLeuLysSerPheAlaLysCysSerArgLeuAepLeuIlePheGlyLeu 181
Qy      627 AATGCGTTATTAAGAACAGACAGATTTGACGTGAAACAGTTCATATGCTGATGCTCTG 686
Db      182 AsnAlaLeuLeuArgThrProAepLeuArgTrpAenSerSerAenAlaGlnLeuLeu 201
Qy      687 GACTACTGCTCTTCCGAAGGGTATTAATCTTTGGAACTTAGGCAATGAACCTTAACAGT 746
Db      202 AsnTyrCysSerSerLysGlyTyrAenIleSerTrpGlnLeuGlnAenGlnProAenSer 221
Qy      747 TTCCTTAAGAGAGCTGATATTATTTTCATCAATGGGTCGACATTAAGAGAGATTAATTC 806
Db      222 PheTrpLysLysAlaGlnIleSerIleAepGlyLeuGlnLeuGlnGlnAepValGln 241
Qy      807 TTCGATTAATCTTAAAGAAAGTCCACCTTCMAAATGCAAACTATAGTCTGTGATGTT 866
Db      242 LeuHisLysLeuLeuGlnIlySerSerAlaPheGlnAenAlaLysLeuLysTrpProAep 261
Qy      867 GGTCAAGCTTCGAAAGAGAGCGCTAAGATGCTGAAGAGCTTCGAAAGCTGTGAGAGAG 926
Db      262 GlyGlnProArgLysThrValLysLeuLeuAepSerPheLeuLysAlaGlyGln 281
Qy      927 GTGATTAATCAAGTATGATGCTGATCTATTTTAATGAGAGAGCTTACAGAGAGAG 986
Db      282 ValIleAepSerLeuThrTrpHisLysTyrLysAenGlyArgValAlaThrLysGln 301
Qy      987 GATTTTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1046
Db      302 AspPheLeuSerSerAepValLeuAepThrPheIleLeuSerValGlnIlySleuLys 321
Qy      1047 GTGGTTGAGAGCACGAGGCTGCGAAGAGGCTGTTAGAGAGAACAGCTTGTGATAT 1106

```

```

Db      322 ValThrLysGluMetThrProGlyLysValTrpLeuGlyGlnThrSerSerAlaTyr 341
QY      1107 GGAGGCGAGCGCCCTGTATCCGACACTTGTGAGCGGCTTATGNGCGGTAA 1166
Db      342 GlyGlyGlyAlaProLeuSerSerThrPheAlaAlaGlyPheMetTrpLeuSplys 361
QY      1167 TTGGGCTGTGACCCGGAATGGAAATGAAGTGGTATGAGCAAGTATCTTTGGAGA 1226
Db      362 LeuGlyLeuSerAlaGlnLeuGlyLeuGlyValValMetArgGlnValPheNeglyAla 381
QY      1227 GGAATCACTACATTATGATGATGAAATCTTCATCTTATGATTTATGCTATCTT 1286
Db      382 GlyAlaThrLysLeuValAspGlnAsnProLeuProAspTyrTrpLeuSerLeu 401
QY      1287 CTGTCAGAAATTTGGTGGACCAAGAGTGTATATGCAAGCGTGGCAAGTTCAAAGAGA 1346
Db      402 LeuPheLysLeuValGlyProLysValLeuMetSerArgValLysGlyProAspArg 421
QY      1347 AGAAGCTTCAGATATACCTTATGACAAACACTGACATCCAGATTAAGTAAAGAGA 1406
Db      422 SerLysLeuArgValTyrLeuHisCysThrAsnValTyrHisProArgTyrArgGly 441
QY      1407 GATTTAATCTGTATGCCATAAATCTCCATAACGTACCAAGTACTTGGCGTTACCTTAT 1466
Db      442 AspLeuThrLeuTyrValLeuAsnLeuHisAsnValThrLysHisLeuLysLeuProPro 461
QY      1467 CTTTTCATCAAGACAGTGAATTAATACCTTCTTAAGACCTTTGGAGCTCATGATTA 1526
Db      462 ProMetPheSerArgProValAspLysTyrLeuLysLeuPheLysSerAspGlyLeu 481
QY      1527 CTTTCCAAATCTGTCCAACTCAATGCTTAAAGTGTGATGATGATCAAACTTGG 1586
Db      482 LeuSerLysSerValGlnLeuAsnGlyGlnThrLeuLysMetValAspGlnGlnThrLeu 501
QY      1587 CCACCTTATATGAAATACCTTCGCGCAAGAGTTCACCTGCGCTTGGCACTTCTCA 1646
Db      502 ProAlaLeuThrGlnLysProLeuProAlaGlySerSerLeuSerValProAlaPheSer 521
QY      1647 TTTAGTTTTTGTGATTAAGAAATGCCAAAGTTGCTGCTTGATC 1691
Db      522 TyrGlyPhePheValIleArgAsnAlaLysIleAlaAlaCysIle 536

RESULT 6
Q90YK5 PRELIMINARY: PRT: 523 AA.
ID      Q90YK5;
AC      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Heparanase.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxId=9031;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      PubMed=11387326;
RA      Goldshmidt O., Zcharia E., Aingorn H., Gatta-Rangini Z., Atzmon R.,
RA      Michal I., Pecker I., Mitrani B., Vlodavsky I.;
RT      "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT      Determined by Their Signal Peptide Sequence.";
RL      J. Biol. Chem. 276:29178-29187(2001).
DR      EMBL, AT037007, AK82648.1, -.
DR      InterPro, IPR005199, Glyco_hydro_79N.
DR      Pfam, PF03662, Glyco_hydro_79n, 1.
SQ      SEQUENCE 523 AA: 58386 MW: 8880B7B18C9BF81 CRC64;

```

```

Query Match: 52.76% Indels: 11
DB: 13 Gaps: 3
us-10-676-079-3 (1-1721) x Q90YK5 (1-523)

QY      99 CTGATGCTGTCTCTCTGAGGCGGCTGGGTCCTCCCTCCCTGGCGCCCTGCCAGCT 158
Db      2 LeuValLeuLeuLeuValLeuLeuLeuAlaValProPro----- 15
QY      159 GCGCAAGACAGAGAGTGTGAGACTTGACTTTCACCCAGAGACCGCTGCACCTGGTG 218
Db      16 -----ArgArgThrAlaGlnLeuGlnLeuGlyLeuArgGlnProIleGlyAlaVal 32
QY      219 AGCCCTGCTCTCTCTCCGTCACATGACGCACTGACGCAACGCGGCGGCTCTC 278
Db      33 SerProAlaPheLeuSerLeuThrLeuAspAlaSerLeuAlaArgAspProArgPheVal 52
QY      279 ATCTCTGGGTTCTCCAAAGCTTGATCTTGAGCAAGGCTTGCTCTGCGTACCTG 338
Db      53 AlaLeuLeuArgHisProLysLeuHisThrLeuAlaSerGlyLeuSerProGlyPheLeu 72
QY      339 AGGTTTGGTGGACCAAGACAGACTTCTTAATTTTCGATCCCAAGAGATCAACTTT 398
Db      73 ArgPheGlyGlyThrSerThrAspPheLeuIlePheAsnProAsnLysAspSerThrTrp 92
QY      399 GAAAGAGAGATTTACTGGCAATCTCAAGTCAACCAAGATTTTGGAAATATGATTCATC 458
Db      93 GlnGlnLysValLeuSerGlnPheGlnAla---LysAspValCysGlnAlaTrpProSer 111
QY      459 CCTCTGATGTGGAGAGAGATTAAGTTCGATGATGAGCCCTACCAAGACAAATGCTACTC 518
Db      112 PheAlaValAlaProLysLeuLeuThrGlnTrpProLeuGlnGlnLysLeuLeu 131
QY      519 CGAAGACACTACCAAAAAATTCAAGAACAGACACTTCAAGAAAGCTCTGTATATGTG 578
Db      132 AlaGlnHisSerTrpLysHisLysAsnThrTrpIleThrArgSerThrLeuSplyle 151
QY      579 CTATACCTTTTGGCAACGCTGCAGACTGACCTGATCTTGGCTTAATCGTTATTA 638
Db      152 LeuHisThrPheAlaSerSerSerGlyPheArgLysValPheGlyLeuAsnAlaLeuLeu 171
QY      639 AGAAGACAGATTTTCAAGTGAACAGTTCTATGCTCAAGTTCCTCGAGTACTGCTCT 698
Db      172 ArgArgGlnGlyLeuGlnTrpAspSerSerAsnAlaLysGlnLeuGlyTyrCysAla 191
QY      699 TCAGAGGGGTATTAATTTCTTGGGAATAGGCAATGAACCTTAACAGTTTCTTAAAGAG 758
Db      192 GlnArgSerTyrAsnIleSerTrpGlnLeuGlnProAsnSerPheArgLysLys 211
QY      759 GCTGATATTTTCATGAATGGGTCGAGTTCAGAGAAAGATTAATTCATTCGATTAACCT 818
Db      212 SerGlyIleCysIleAspGlyPheGlnLeuGlyArgAspPheValHisLeuArgGlnLeu 231
QY      819 CTA---AGAAAGTCCACCTTCAAAATGCAAAATCTATGCTCGATGTTGGTCAAGCT 875
Db      232 LeuSerGlnHisProLeuTyrArgHisAlaGlnLeuTyrGlyLeuAspValGlyGlnPro 251
QY      876 CGAAGAAAGACGCTTAAGATGCTGAAGACTTCTGAAAGCTGTGGAGAACTGATGAT 935
Db      252 ArgLysHisLysThrGlnHisLeuLeuArgSerPheMetLysSerGlyGlyLysAlaIleAsp 271
QY      936 TCAAGTTACATGACACACTCATTTTGAATGAGACGACCTGCAACGAGAGATTTCTGA 995
Db      272 SerValThrTrpHisHisLysTyrValAsnGlyArgSerAlaThrArgGlnAspPheLeu 291
QY      996 AACCTGATGTATGACATTTTATTCATCTGTGCAAAAGTTTCCAGTGGTGTAG 1055
Db      292 SerProGlnValLeuAspSerPheAlaThrAlaIleHisAspValLeuGlyIleValGln 311
QY      1056 AGCACCAAGGCTGGCAAGAGGCTGTGTTAGAGAAACAAGCTTGCAATATGAGGCGGA 1115
Db      312 AlaThrValProGlyLysLeuValTrpLeuGlyGlnThrGlySerAlaTyrGlyGly 331

```



QY	1116	GGGCGCTTGCTATCCGACACCTTGTGACCTGGCTTATGAGCGGAGTAATTAATTTGGGCTG	1175
Db	332	AlaProGlnLeuSerAsnThrTyrValAlaGlyPheMetTrpLeuAspGlyLeu	351
QY	1176	TCAGCCGAAATGGAAATGAAAGTGCTGATGAGCGCAATGATCTTTGGAGCGAAACTAC	1235
Db	352	AlaIlaaArgAaGrgYlIleAspValValMetCArgGlnValSerPheGlyAlaGlySerTyr	371
QY	1236	CATTATGAGTGAAGTGAATACTTGATCCTTTACTGTATATATGAGTATCTCTGTCTCAAG	1295
Db	372	HisLeuValaIleAspAlaGlyPheLeuProLeuProAspTyrTrpLeuSerLeuLeuTyrIys	391
QY	1396	AAATTTGTGGCGACCCAGCTGTGTTATGCGCAAGCTGCGCAAGGTTCAAAAGAGAAAGCTT	1355
Db	392	ArgLeuValaGlyThrArgValIleuGlnIlaSerValaGluIlaAspAlaArgArgPro	411
QY	1356	CGAGTATACCTTTCATTTGACCAAAACATGTACATCCAGCTTTAAAGAGAGATTAACT	1415
Db	412	ArgValTyrLeuHisAsCysThrAsnProArgHisProTyrIysTyrArgGluGlyIaAspValaThr	431
QY	1416	CTGTATGCGCATTAACCTCTCATACGTCAACCAAGTCTTGGCGGTTAACCTTATCCTTTTCT	1475
Db	432	LeuPheIaIleAsnLeuSerAsnValaThrGlnSerLeuGlnLeuProIysGlnLeuTyr	451
QY	1476	AACACGACAGTGAATAAATACCTTTCTTAAGACCTTTGGAGCCTCATGATTACTTTCCAA	1535
Db	452	SerIysSerValaIleArgGlnTyrLeuLeuLeuProHisGlyIysAspSerIleLeuSerArg	471
QY	1536	TCGTGCGCAACCTCATGCTCTAACTCTTAAGTGGTGAATGATATCAACCTTGGCACCTTTA	1595
Db	472	GluValaGlnLeuAsnGlyIaArgLeuLeuGlnMetValaAspAspGluThrLeuProIaLeu	491
QY	1596	ATGGAAAAACCTCTCCGGCGCAAGAGTTCACTGGGCTTGCACGTTTCTCATATAGTTT	1655
Db	492	HisGluMetAlaIleuAlaProGlySerThrLeuGlyLeuProIaPheSerTyrGlyPhe	511
QY	1656	TTTGATGATGAAGAAATGCCAAAGTTCCTGCTTGCAATC	1691
Db	512	TyrValIleIleArgAsnAlaIaIleAlaIleAlaCysIle	523

RESULT 7

Q9HB37 PRELIMINARY PRT 592 AA.

AC Q9HB37 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Heparanase-like protein HPA2c.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=20483645; PubMed=11027606;

RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,

RL Hiccock M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.,

RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian Heparanase Family Member.";

RT Biochem. Biophys. Res. Commun. 276:1170-1177(2000).

RL EMBL; AF282887; AAC34423.1; -

DR GO; GO:0005622; C:intracellular; TAS.

DR GO; GO:0030305; F:heparanase activity; TAS.

DR InterPro; IPR005199; Glyco\_hydro\_79N.

DR Pfam; PF03662; Glyco\_hydro\_79n\_1.

DQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;

Alignment Scores:	
Pred. No.:	4.67e-85
Score:	1154.50
Percent Similarity:	57.84%
Best Local Similarity:	43.55%
Query Match:	37.02%
Length:	592
Matches:	250
Conservative:	82
Mismatches:	189
Indels:	53

DB:	4	Gaps:	9
US-10-676-079-3 (1-1721) x Q9HB37 (1-592)			
OY	120	CCGCTGGTCCCTCTCTCCCTGGCGCCCTG-----	149
DB	18	ProProAlaCybSeuAlaProGlyAlaLeuThyLeuAlaLeuLeuHisLeuSerLeu	37
OY	150	-----CCCCGACCTGGCG-----	CAAGCA 167
DB	38	SerSerGlnAlaGlyAerAraGlyAraGlyProLeuProValAlaAraGlyAlaGlyLeuValGlyLeu	57
OY	168	CAGGACGTCGTGGACCTGGACCTTTTTCACCCAGAGACGCTGGACCTGTAGCCCTTCG	227
DB	58	LyserThreuleuLeuLeuAerValSerThryLeuAerProValAraGThryValaGlnAra	77
OY	228	TTCTCTGTCCGTACACCATTTGACGCCAACCTGGCCACGAGACCCGGGTTCTCATCTCTCTG	287
DB	78	PhenLeuSerLeuGlnLeuAerProSerLeuAlaHisAer--GlyThryLeuAerPhenLeu	96
OY	288	GGTTCTCCAAAGCTTCTGACCTTGGCCAGAGAGCTGTCTCTCTGCTGACGATTTGGT	347
DB	97	SerSerLyserAraGlyValAlaThreAlaAlaArgGlyLeuSerProAlaPhenLeuAraGlyPheGly	116
OY	348	GGACACAGACAGACACTTCTTAATTTTC-----GATCCCAAGAAAGAAATCAAC	395
DB	117	GlyLyserAraGThryAerPhenLeuGlnPheGlnAlaSerLeuAraGlyAerProAlaLyserAraGly	136
OY	396	TTTGAAGAGAGAGATTACTGGCAATCTCAAGTCAACCGAATATT-----	440
DB	137	GlyProGlyProAerLytyTyLeuLybAerTyGlyAerAerAlaValAraGlySerAerVal	156
OY	441	-----TGCAAATATGATTCATCCCTCTCTGATGTGGAGAGAG	479
DB	157	AlaLeuAerLyserGlnLyGlyCyLeuAlaAlaGln--HisProAerValaMetLeuGln	175
OY	480	TTAGCGTTGGAAATGGCCCTACACAGAGCAATTG---TACTCCGAGAAACACTACCGAA	536
DB	176	LeuGlnAraGlyValaAlaAlaGlnMetHisLeuValaLeuLeuLyserGlnGlnPheSerAer	195
OY	537	AAGTTTCAAGAACAGACCTACTACCAAGAGCTGTAGATGTGTATCACTTTTGGCAAC	596
DB	196	ThrySerAerLeuLeuLeuLeuThryAlaArgSerLeuAerLyserLytyAerPheAlaAer	215
OY	597	TGCTCAGACTGGACCTTGATCTTTTGGCCCTAAATGCGTTATTAAAGACAGCAATTGGAC	656
DB	216	CysSerGlyLeuHisLeuAlaPheAlaLeuAerAlaLeuAerAraGlyAerProAerAerSer	235
OY	657	TGGAACAGTTTAAATGCTACAGTTGCTCTGTGACACTGCTCTTCCAAAGGGGTAAACATT	716
DB	236	TThpAerSerSerAlaLeuSerLeuLeuLytySerAlaSerLybLytyTyAerAla	255
OY	717	TCTTGGGAATGAGCAATGAACTCAACAGTTTCTTAAAGAGGCTGATATTTTCACTAAT	776
DB	256	SerThryGlnLeuGlnAraGlnProAerAerValAraGThryMetHisGlyAraGlyAlaValaAer	275
OY	777	GGGTGCGAGTTTGGAGAAAGATTATATCACTTGCATTAACCTGTAAAGAG---TCCACC	833
DB	276	GlySerGlnLeuGlyLybAerLytyLeGlnLeuLySerLeuLeuGlnProIleAerGlyLe	295
OY	834	TTTCAAAATATGCAAAACCTATGCTGCTGCAAGTTGGTGTGACGCTCGAAGAAAGACGGCTAAG	893
DB	296	TySerAerAraGlyAerSerLeuTyGlyProAerAlaGlyAraGThryAerAerValaIleAla	315
OY	894	ATGCTTGAAGAGCTTCTGAAAGCGCTGGTGGAGAAATGATTTATTCAGTTTACATGGCATCAC	953
DB	316	LeuLeuAerAerGlyPheMetLybValaAlaGlySerThryValaAerAlaValaThryTThpGlnHis	335
OY	954	TACATATTTGATGAGACGACGCTGACACGAGGAGATTTTGTAAACCTGATGTATTGGAC	1013
DB	336	CysThryTlleAerAraGlyAraGlyValaValaValaMetAerPhenLeuLythryAraGlyLeuAer	355
OY	1014	ATTTTATTTTCATCTGTGCAAAAAGTTTTCAGAGTGGTTGAGAGCACAGGCTGGCAG	1073



```

Db      356 ThrLeuSerAspGlnIleArgLysIleGlnLysValValAsnThrTyrThrProGlyLys 375
QY      1074 AAGGCTGGTGGAGAAACAAGCTCTGATATGAGAGCGCCCTTGCTATCCGAC 1133
Db      376 LysIleTyrPheGlnGlyValValThrThrSerAlaGlyGlyThrAsnAsnSerAsp 395
QY      1134 ACCCTTGACAGCTGGCTTTATGTGGTGAATTAATGGCGCTGACGCCGAATGGGAATA 1193
Db      396 SerTyrAlaIleGlyPheLeuTyrPheLeuAsnThrLeuGlyMetLeuAlaAsnGlnGlyIle 415
QY      1194 GAAAGTGTGATGAGCAAGCAATATTCTTTGAGCAAGAAATCAATCAATTAGTGAATAAAC 1253
Db      416 AspValIleAlaIleArgHisSerPhePheAspHisGlyTyrAsnHisIleValAspGlnAsn 435
QY      1254 TTGCAATCCTTACCTGATATGCTGATATCTCTTCTGTTCAAGAAATGGTGGCAACGAC 1313
Db      436 PheAsnProLeuProAspTyrTyrPheSerLeuLeuTyrLysArgLeuIleGlyProLys 455
QY      1314 GTGTTAATGGCAAGCGGTGCAAGGTTCAAGAGAACG----- 1349
Db      456 ValLeuAlaValHisValAlaIleGlnArgLysProArgProGlyArgValIleArg 475
QY      1350 ---AAGCTTCCAGATATCTTCAATGACAAACATGACATCAAGTAAAGTAAAGAGA 1406
Db      476 AspLysLeuAlaArgIleTyrAlaHisCysThrAsnHisIleAsnHisValAlaArgGly 495
QY      1407 GATTAACTCTGTATGSCCAATCAACCTCCATTAACGTCAACACATGACTGCGGTACCTCAT 1466
Db      496 SerIleThrLeuPheIleIleAsnLeuHisArgSerArgLysLysIleLysLeuAlaGly 515
QY      1467 CTTTTTCTTAACAGCAAGTGAATGAATATCTTCTTAAGACCTTTGGACCTGATGATTA 1526
Db      516 ThrLeuArgAspLysLeuValHisGlnTyrLeuLeuGlnProTyrGlyGlnGlyLeu 535
QY      1527 CTTTCCAAATCTGTCCAATCAATGCTTCAACTCTTAAGATGGTGGATGATCAACCTTG 1586
Db      536 LysSerLysSerValGlnLeuAsnGlnGlnProLeuValMetValAspAspGlyThrLeu 555
QY      1587 CCACCTTTTAATGGAATAAACTCTCCGCGCCAGCAAGTCACTGGGCTTGCCACTTTCTCA 1646
Db      556 ProGlnLeuLysProArgProLeuArgAlaGlyArgThrLeuValIleProProValThr 575
QY      1647 TATAGTTTCTTGATAGAATAAGTCCAAAGTTGCTGCTGC 1688
Db      576 MetGlyPhePheValValLysAsnValAlaMetAlaLeuAlaCys 589

RESULT 8
Q8WMQ2 PRELIMINARY; PRT; 592 AA.
AC Q8WMQ2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DB Heparinasae 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Peesegue Safontas B.J.O.P.S.;
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ299719; CAC82491.1; -
DR Genew; HGNC:18374; HPS2.
DR Interpro; IPR005199; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.

```

```

SQ SEQUENCE 592 AA; 66520 MW; 9478841FEACD558B CRC64;
Alignment Scores:
Pred. No.: 2,096-84 Length: 592
Score: 1146.50 Matches: 249
Percent Similarity: 57.67% Conservative: 82
Best Local Similarity: 43.38% Mismatches: 190
Query Match: 36.76% Indels: 53
DB: 4 Gaps: 9

US-10-676-079-3 (1-1721) x Q8WMQ2 (1-592)
QY 120 CCGTGGGTCCTCTCCCTGGCGCCCTG----- 149
Db 18 ProProAlaCysLeuAlaProGlyAlaLeuTyrLeuAlaLeuLeuHisIleuSerLeu 37
QY 150 -----CCCGACCTGCG-----CAAGCA 167
Db 38 SerSerGlnAlaGlyAspArgArgProLeuProValAspArgAlaAlaGlyLeuLysGlu 57
QY 168 CAGGACGTGCGACCTGGACCTTCTTCAACCCAGAGCCGCTGACCTGAGCCCTCG 227
Db 58 LysThrLeuIleLeuLeuAspValSerThrLysAsnProValArgThrValAsnGlyAsn 77
QY 228 TTCTGTCTCCATCCATTGACGCGCAACCTGGCCAGCAGACCCGCGGTCTCATCTCTG 287
Db 78 PheLeuSerLeuGlnLeuAspProSerIleIleHisAsp---GlyTyrPheAspPheLeu 96
QY 288 GGTCTTCCAAAGCTTGGATCTTGGCCAGAGGCTTGTCTCTGCTGATCACTGAGTTTGGT 347
Db 97 SerSerLysArgLeuValThrLeuAlaArgGlyLeuSerProAlaPheLeuArgPheGly 116
QY 348 GGCACCAAGACAGACTTCTCAATTTT-----GATCCCAAGAGGAATCAAC 395
Db 117 GlyLysArgGlnAspPheLeuGlnPheGlnAsnLeuAlaGlnAsnProAlaLysSerArgGly 136
QY 396 TTTGAAGAGAGAAGTTACTGGCAATCTCAAGTCAACAGCATATT----- 440
Db 137 GlyProGlyProAspTyrTyrLeuLysAsnTyrGlnAspAspIleValArgSerAspVal 156
QY 441 -----TGCAATATGAGATTCATCTCTGATGTGGAGAGAG 479
Db 157 AlaLeuAspLysGlnLysGlyCysLysIleAlaGln---HisProAspValMetLeuGlu 175
QY 480 TTACGCTTGGATGAGCCCTTACCAGAGCAATTG---CTACCTCCAGACACTACAGCAAA 536
Db 176 LeuGlnArgGlyLysAlaAlaGlnMetHisIleValLeuLysGlnGlnPheSerAsn 195
QY 537 AAGTTCAGAGACAGCACTTCAAGAAAGCTGTGATGATGCTATACACTTTTCAAC 596
Db 196 ThrTyrSerAsnLeuIleLeuThrAlaArgSerLeuAspLysLeuTyrThrAsnSerAlaAsp 215
QY 597 TGTCCAGAGACTGAGACTTATGCTTTGGCTTAATGCGTTATTAAGAACAGCAATTGGAG 656
Db 216 CysSerGlyLeuHisIleLeuIlePheAlaLeuAsnAlaLeuAlaGlyArgAsnProAsnAsnSer 235
QY 657 TGGACAGCTTCAAGTCAAGTGTCTGCTGAGACTCTGCTTCCAAAGGGATTAACATT 716
Db 236 TrpAsnSerSerSerSerLeuSerLeuSerLeuLysTyrSerAlaSerLysLysTyrAsnIle 255
QY 717 TCTTGGAGACTAGGACATGAACCTTAACAGTTTCTTAAAGAGCTGATATTTCATCAT 776
Db 256 SerTyrGlnLeuGlyAsnGlnProAsnAsnTyrArgThrMetHisGlyArgAlaValAsn 275
QY 777 GGGTCCAGTTAGAGAAAGATTATTTCAATTGCAATAACTTTAAGAAAG---TCCACC 833
Db 276 GlySerGlnLeuGlyLysAspTyrIleGlnLeuLysSerLeuLeuGlnProIleArgIle 295
QY 834 TTCAAAATATGCAAAATCTATGCTGCTGATGCTGCTGCTGCAAGAAAGACGCTAAG 893
Db 296 TyrSerArgAlaSerLeuTyrGlyProAsnIleGlyArgProGlyLysAsnValIleAla 315
QY 894 ATGCTGAAGAGCTTCTGAAGGCTGTGAGAGAAAGTATGATTGATTCAATTACATGCGATCAC 953

```

```

Db      316 LeuLeuaspGlyPheMetLysValAlaGlySerThrValAspAlaValThrProGlnHis 335
QY      954 TACTATTGATGAGCCGAGCTGTACACAGGAGATTTCTTAACCTCGATGATTTGAC 1013
Db      336 CyetYrLleAspGlyArgValValLysValMetAspPheLeuSerThrArgLeuLeuAsp 355
QY      1014 ATTTTATTTTCATCTGTGCAAAAAGTTTCCAGGTGTGAGACACAGCCGTGGCAG 1073
Db      356 ThrLeuSerAspGlnLleArgLysIleGlnValValAsnThrValProGlyLys 375
QY      1074 AGGTCTGTAGGAGAAACAGCTGTGATGAGAGCGGAGCCCTGTATCCGAC 1133
Db      376 LysIleTrpLeuGlnGlyValValThrThrSerAlaGlyGlnTrpAsnLeuSerAsp 395
QY      1134 ACCTTTCAGCTGTGCTTTATGTGGCTGATTAATTTGGCCCTGTGACCCGATGGGAATA 1193
Db      396 SerThrAlaAlaGlyPheLeuTrpLeuAsnThrLeuGlyMetLeuAlaAsnGlnGlyIle 415
QY      1194 GAAGTGTGATGAGGCAAGTATTTCTTGGACAGGAACTACCTTTAGTGTGATAAAC 1253
Db      416 AspValValIleArgHisSerPheAspPheGlyTrpAsnHisLeuValAspGlnAsn 435
QY      1254 TTGGATCCTTACCTGATTTATTTGGCTATCTCTGTGCAAGAAATTTGGGGGACCAAG 1313
Db      436 PheAsnProLeuProAspGlyTrpLeuSerLeuLeuGlyTrpArgLeuIleGlyProLys 455
QY      1314 GTGTAAATGCGAACGCTGCAAGGTTTCAAGAGAGG----- 1349
Db      456 ValLeuAlaValHisValAlaGlyLeuGlnArgLysProArgProGlyArgValIleArg 475
QY      1350 ---AGCTTGTGATATCTCTCATTTGCAAAACATGACATTCAGATTAAGAAAGA 1406
Db      476 AspLysLeuArgLysIleGlyTrpAlaHisCysThrAsnHisAsnIleAsnValArgGly 495
QY      1407 GATTTAACCTGTATGTCATTAACCTTCATAGCTCAACAGTATCTTGGCGGTATACCTAT 1466
Db      496 SerIleThrLeuPheIleIleLeuLeuHisArgSerArgLysValIleLysLeuAlaGly 515
QY      1467 CTTTTTCTAACAACAGATGATTAATACCTTCTTAAGACCTTGGAGCTCATGATTA 1526
Db      516 ThrLeuArgAspLysLeuValHisGlnTrpLeuLeuGlnProGlyGlnGlnGlyLeu 535
QY      1527 CTTTCCAAATCTGTCCCACTCAATGTCTTAACCTTAAGATGTGGATGATCAAACTTG 1586
Db      536 LysSerLysSerValGlnLeuAsnGlyGlnProLeuValMetValAspAspGlyThrLeu 555
QY      1587 CCACTTTAATGGAACCACTCTCCGCGCAAGAACTTCACTGGGCTTGCAGCTTCTCA 1646
Db      556 ProGlnLeuLysProArgProLeuArgAlaGlyArgThrLeuValIleProValThr 575
QY      1647 TATAGTTTTTTGTGATMAAATGCAAAAGTTGCTTGC 1688
Db      576 MetGlyPhePheValValLysAsnValAlaAlaLeuAlaLys 589

RESULT 9
Q8WMQ1 PRELIMINARY; PRT; 548 AA.
AC Q8WMQ1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-UN-2003 (TrEMBLrel. 24, Last annotation update)
DE Heparanease 3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Peasegue Safoncas B.J.O.P.S.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Legoux P., Legoux R., O'Brien D., Salome M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ299720; CAC82492.1; -
DR InterPro; IPR005139; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79n; I.
SQ SEQUENCE 548 AA; 61771 MW; B8986303FC73A60A CRC64;

Alignment Scores:
Pred. No.: 9, 08e-74 Length: 548
Score: 1015.50 Matches: 224
Percent Similarity: 56.53% Conservative: 79
Best Local Similarity: 41.79% Mismatches: 180
Query Match: 32.56% Indels: 53
DB: Gaps: 9

US-10-676-079-3 (1-1721) x Q8WMQ1 (1-548)
QY 84 GCGCTGCCGCCGCGCTGATGCTGTCTGCGGCGCGCTGCCCTTCTCCCTGCG 143
Db 6 AlaPheProGlnAlaMetLeuSerSerAsnSerArgProProAlaCysLeuAlaProGly 25
QY 144 GCCCTG----- 149
Db 26 AlaLeuTrpLeuAlaLeuLeuLeuHisLeuSerLeuSerSerGlnAlaGlyAspArg 45
QY 150 CCCCAGCTGCG-----CAAGCAGAGAGCTGCGTGAAGCTTGC 191
Db 46 ProLeuProValAspArgAlaAlaGlyLeuLysGlnLysThrLeuIleLeuLeuAspVal 65
QY 192 TTCACCCAGAGCGCTGCACTGCTGTGAGCCCTGCTCTGCTGCTGCAATGAGCC 251
Db 66 SerThrLysAsnProValArgThrValAsnGlnAspPheLeuSerLeuGlnLeuAspPro 85
QY 252 AACCTGCGCAGGAGCCGCGGCTTCATCCCTCGTGGTTCCAAAGCTTGACCTTG 311
Db 86 SerIleIleHisAsp---GlyTrpLeuAspPheLeuSerSerLysArgLeuValThrLeu 104
QY 312 GCCAGAGGCTGTCTCTGCTGATGCTGAGATTTGGTGGCAGCAGACATTTCTTAAT 371
Db 105 AlaArgGlyLeuSerProAlaPheLeuArgPheGlyGlyLysArgHisPheLeuGln 124
QY 372 TTC-----GATCCCAAGAAAGAAATCACTTGAAGAGAAATTAAGTGGCAA 419
Db 125 PheGlnAsnLeuArgAsnProAlaLysSerArgGlyGlyProGlyProAspGlyTrpLeu 144
QY 420 TTCAGATCAACAGATTAAT-----TGC 443
Db 145 LysAsnTrpGlnAspAspIleValAlaArgSerAspValAlaLeuAspLysGlnLysGlyCys 164
QY 444 AATATGATTCATCCCTCTGATGTGAGAGAAAGTTACGTTGGAATGGCCCTTACAG 503
Db 165 LysIleAlaGln---HisProAspValMetLeuGlnLeuGlnArgLysValAlaAlaGln 183
QY 504 GAGCAATTG---CTACTCCGAGAACTACAGAAAGATTCAAGACAGCACTTACTCA 560
Db 184 MetHisLeuValLeuLeuLysGlnGlnPheSerAsnThrTyrSerAsnLeuLeuLeu 203
QY 561 AGAAGCTGTGATGATGCTATACATTTTGCAAACTGTGCGAGATGCACTTGATCTTT 620
Db 204 AlaArgSerLeuAspLysLysLeuTrpAsnSerAlaAspCysSerGlyLeuHisLeuIlePhe 223
QY 621 GGCCTTAATGCTTATTAAGAACAGACAGATTTGAGTGAAGACATTAATGCTCAGTTG 680
Db 224 AlaLeuAsnAlaLeuArgHisProAsnAsnSerTrpAsnSerSerHisLeuSer 243
QY 681 CTCCTGACTACTGCTCTTCCAAAGGGGATTAACATTTCTTGGAACTAGCAATGAACT 740
Db 244 LeuLeuLysTrpSerLysSerLysLysValTrpAsnIleSerTrpGlnLeuGlnGlnGlyPro 263
QY 741 AACAGTTCTTAAAGAGCTGATATTTTCATCAATGAGGTTCCAGTTAGAGAAATTTAT 800

```

Db	264	AmbientlyRgAgLThmMehni6G1yAkgb1aValaMbngUysertG1nleuG1yLyaA8prYr	283
Qy	801	ATTCAGTTCGACTAACCTTCTTAAGAAAG--TCCAGCTTCAGAAATGCGAAACTSTAGST	857
Db	284	l1eglnleuUysertleuGlnPro1leArg1leTyRserAArg1aSerleuYrG1y	303
Qy	858	CTGATGTTGGTACGAGCTCGAAGAAAGCGGCTAAGATGCTGAAGAGCTTCCTGAAGCT	917
Db	304	Pro5n1l1eg1yAArgPro4yRyLysanVal11ealaleuLeuA8prG1yPhmeUysVal	323
Qy	918	GGTGGAGAGATGATGATTCATTCAGTTACATGAGCATCACTATTCATTAAGATGAGCGACTCT	977
Db	324	AlaG1ysertRvAlaA8prAlaVal1ThrTrgln1n1eGyRy1leA8prG1yAArgVala1	343
Qy	978	ACCAAGGAGAGATTTCTTAACCCCTGATGTATTCAGCATTTTATTCATCTGTGCAAAA	1037
Db	344	LysVal1MeA8prPhelUys1eThRArgUleuA8prThRLeuSerA8prG1n1leArg1n	363
Qy	1038	GTTTTCACAGGAGTGTAGAGACACCGAGCTCGCAAGAGGTCGTGTGAGCAAAACAGC	1097
Db	364	l1eglnUysValaVala8nthRyThRProG1yLys1le1eRpleGlnG1yVala1	383
Qy	1098	CTTCGATGTAGAGCGGAGCGCCCTGTGATCTCGACACCTTGTGAGCTGGCTTATGTGG	1157
Db	384	ThRThrserr1aG1yG1yThRAnbnLeuSerA8prTyR1a1a1G1yPhelUeUtr	403
Qy	1158	CTGATTAATTTGGCGCTGTGACCCCGAATGGGATAGATGAAGTGATGAGCGCAAGTATTC	1217
Db	404	Leu8nthRleuG1yMeUleuAlaMbng1nG1y1leA8prVala11leA8rH1serRhe	423
Qy	1218	TTTGGAGAGGAAACTATTCATTTAGTGATGAAAACTTCGATCCTTTACCTGATTAATGG	1277
Db	424	Ph88prH1sG1yTyRAnbn1leUvalA8prGlnA8prA8prLeuA8prTyR	443
Qy	1278	CTATCTCTTCGTGCAAGAAATGTGTGGGACCAAGGTTATATGCGAAGCGCAAGCT	1337
Db	444	Leu8erUleUyRyLysArgLeu1leG1yProUysVal1leuAlaVala1n1eVala1aG1y	463
Qy	1338	TCAAGAGAGAG-----AAGCTTCGATATACCTTCAT	1370
Db	464	LeuGlnAArgUysPro4yRProG1yAArgVal11leA8prA8prUysleuA8rG1leTyRAlaH1s	483
Qy	1371	TGCGAACAACATGACATATCCAAAGGTATTAAGAAGAGATTTATCTGTATGCCATAAC	1430
Db	484	CysThRAnbn1bH1sAnbn1sA8nTyRValAArgG1ysert1eThRleuPhel1e1eA8n	503
Qy	1431	CTCCATTAACGTCACCAAGTATCTGGGGTATCCCTATCCTTTTTCACACAGAGTGGAT	1490
Db	504	LeuH1sA8prA8prUys1eUys1eUysleuUalG1yThRleuA8prUysleuValH1s	523
Qy	1491	AAATACCTTCTTAAGACCTTTGGGACCTCATGGATTAATCTTCCAAATCT	1538
Db	524	GlnTyRleuLeuGlnProTyRGlYcInGlnG1yLeuUysert1yRThr	539
RESULT 10			
Q9HB38			
ID	Q9HB38	PRELIMINARY;	PRT; 534 AA.
AC	Q9HB38;		
DT	01-MAR-2001 (TReMBLrel. 16, Created)		
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)		
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)		
DE	Heparanase-like protein HP42b.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelaeostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_Txid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hitcock M., Patel S., Barry E., Stuberfield C., Territt J., Page M.,		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
	Heparanase Family Member.";		

Bt	Biochem Biophys Res Commun.	276:	1170-1177(2000).
Df	EMBL; AF282286; MAG23422.1; .-		
Df	InferPro; IPR005199; Glyco_hydro_79N.		
Df	Pfam; PF03662; Glyco_hydro_79n; 1.		
SQ	SEQUENCE	534 AA;	60063 MW; C3DESE90CB338CA CRC64;
 Alignment Scores:			
Pred. No.:		2,38e-67	Length: 534
Score:		936.50	Matches: 217
Percent Similarity:		51.39%	Conservative: 78
Best Local Similarity:		37.80%	Mismatches: 168
Query Match:		30.03%	Indels: 111
		4	Gaps: 12
 US-10-676-079-3 (1-1721) x Q9HB38 (1-534)			
Oy	120	CCGCTGGTCCCTCCTCCCTCGGCCCTCG-----	149
Dd	18	ProProAlaCysLeuAlaProIlyAlaLeuIleuHisLeuSerLeu	37
Oy	150	-----CCGCAGCTTCG-----CAACA	167
Dd	38	SeSerGIlnAlaGlYAspArgArgProIeuProValAspArgAlaIlaGlYLeuYsglu	57
Oy	168	CAGACGTGCTGGACTGGACTTTCTTCACCAGAGACGGCTGCACCTGGTGAGCCCTCG	227
Dd	58	IyeThrIleuIleuLeuAspValIseThrIlysAsnProValAaGrThrValAsnGluAsn	77
Oy	228	TTCCTGTCGGTACCATTTAGACGCMAACCGGCACCGACCGCGGATTCCTCATCTGCTG	287
Dd	78	PhenLeuSerLeuIleuAspProSerIleIleHisAsp--GlyTrpLeuAspPheLeu	96
Oy	288	GGTTCTCAAAGCTTCGTAACCTTGGCCAGAGGCTTGTCTCCGCGTACTGAAGTTTGT	347
Dd	97	SeSerIlyArgLeuAlaThrIleuAlaArgIlyLeuSerProAlaPheLeuArgPheGly	116
Oy	348	GGCACAAAGACAGACTTCTTAATTTC-----GATCCCAAGAAGAAATCAAC	355
Dd	117	GIlyAsArgThrAspPheLeuIlnPheGlnAsnLeuArgAsnProAlaIylsSerArgIly	136
Oy	396	TTTTGAAGAGAGAGTAGTACTGGCAATCTCAAGTCAACCAAGATATT-----	440
Dd	137	GIlyProGIlyProAspIryTrIyleuYAsnTyrgIusAspAlIleValArgSerAspVal	156
Oy	441	-----TGCAATAATGATTCATCCCTCCTGATGTGGAGAGAG	479
Dd	157	AlaLeuAspIlySglInylSgIlyCyIsylSIleAlaGln--HisProAspValMetLeuGlu	175
Oy	480	TTCAGTTGAGATGSCCTTACCAAGAGCAATG--CTACTCCAGAAACCTACACAGAAA	536
Dd	176	LeuGIlnArgGIlylulysAlaIlaGlnMetHisLeuValIleuLeuYsgIugIlnPhe--	193
Oy	537	AAGTTCAAGAACGACCTACTCAAGAAAGCTCTGAGATGTGCTATACACTTTTGC AAC	556
Dd	194	-----SerAsnThrIyrser-----	198
Oy	597	TGCTCAGACTGGACTTGATCTTGGCTTAATGCGTTATTAAAGACAGACATTTCAG	656
Dd	199	-----AsnLeuIleu-----	202
Oy	657	TGAAACAGTTTAAGCTCA GTTGTCTCTGCACTACGTCTTCCAAGGGATTAACATT	716
Dd	202	-----	202
Oy	717	TCTTGGAACTAGGCAATGAATCACTTAAGCTTCTTAAAGAGCGTATTTTCATCAAT	776
Dd	203	-----ThrIlnProAsnAsnTyraGrThrMetHisGIlyArgAlaValAsn	217
Oy	777	GGGTCCAGTTAGAGAGAAATATATTTCAATTTGCTAAACTTTCAAGAAAG--TCCACC	833
Dd	218	GIYSerGIlnleuClYlyAspIryTrIleGlnleuYsSerIleuLeuGlnProIleArgIle	237
Oy	834	TTCAAAAATGCAAAACTTATGCTCTGATGTTGTGACCTCGAAGAAAAGACGGCTAAG	893

Db	Accession	Species	Gene	Position
Db	238	Tyrosarctylalaserleutyrglylproanileglylarginylproargylvalasvallealea	257	
Qy	894	ATGCTGAGAGAGCTTCTGTAAGGCTGTGTGAGAAAGTATGATTCATGATTAAGCATC	953	
Db	258	LeuleuabpGLyphenetlulsvalalaglyserthvalabpalevalthrtpglinh	277	
Qy	954	TACTATTTGAAAGGACGAGCTGCTCAAGGAGAAATTTTCTAAACCCGTGATTTATTTGAC	1013	
Db	278	CysTyrXileabpGLyArgvalValLysValMetabpRheleutythrXyleuenub	297	
Qy	1014	ATTTTATTTATCTCTGTGCAAAAGATTTTCCAGGTGTGTGAGACACGAGCTGGCAAG	1073	
Db	238	ThrlseuSerabpGLilearglylsileglnlyvalvalabthrthrythrProLylus	317	
Qy	1074	AAGTCTGGTTTGTGAGAAACAAGCTGTGCAATGTGAGGCGAGCCCTGTGCTATCCAGC	1133	
Db	318	LysIletrpLeuglnGlyValValThrtrSerallaglylYthAsnAsnLeuSerab	337	
Qy	1134	ACCTTTGACGTGCTGTTTATGTGGCTGTGATTAATTTGGCTGTGACCCGAATGGAAAT	1193	
Db	338	SerTyrAlaIlaaglypHeuLeutrpeuabthrLleuglyMetleuAlaasnlnglyle	357	
Qy	1194	GAGGTGGTATGAGCAAGATATCTTTTGGAGAGAAATACATTTATGTGTGATTAAGAC	1253	
Db	358	AspValValIleuXhisSerPhePheabpIleGlyTyrAsnIleLeuValAspIlnb	377	
Qy	1254	TTGCATCTTTTACCTGATTTATTTGGTATCTCTTCTGTGCMAAAATGGTGGACCAAG	1313	
Db	378	PheabpProLeuProabpTyrtrPheusertleuTyrLysArgleuIleLylProLys	397	
Qy	1314	GTGTTTAATGGCAAGGTGCAAGTTCMAAGAGAAAG-----	1348	
Db	398	ValleuAlaValIhisValalaglyLeuGlnArgLysProabpProGlyArgValIlearg	417	
Qy	1350	---AAGCTTCGAGTATACCTTCAATTCACCAAAACCTGCAATCCAAAGTATTAAGAGA	1406	
Db	418	AspLysLeuArgLileTyrAlaIleIserYthrAsnIleAsnIleAsnIleValalargly	437	
Qy	1407	GATTTAACTCTGTATGCCATTAACCTCCATTAACGTCAACAGTACTTGGGTGCTTAT	1466	
Db	438	SerIlethrLeuPheIlelleAsnleuIlsargSerArgLysLileLysleuIleagly	457	
Qy	1467	CCTTTTTCACAAACAGTGTGATTAATACCTTTCAAAGCCTTGGAGCTCATGAGATTA	1526	
Db	458	ThrleuArgabpLysLeuValIhisGlnTyrLeuLeuGlnProTyrGlyGlnIleuLysleu	477	
Qy	1527	CTTTGCAATCTGTGCAATCAATGCTGTACTTTAAAGATGTGTGAGATCAACCTTG	1586	
Db	478	LysSerLysSerValGlnleuAsnGlnInProLeuValMetValAspabpLylThrleu	497	
Qy	1587	CCACCTTTAATGGAAAAACCTTCGGGCGAGAAAGTTCACTGGGCTTGCCAGCTTTTCA	1646	
Db	498	ProGlnleuLysProabpProLeuArgIlaaglyArgIleValIleLeuProValThr	517	
Qy	1647	TATAGTTTTTTTGTGATAAGAAATGCCAAAGTGTGCTGTGC	1688	
Db	518	MetGlyPhePheValValLysAsnValAsnAlaLeuAlaCys	531	

RX	MEDLINE=20483645; PubMed=11027606;
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RB	Hancock M., Patel S., Barry E., Snubberfield C., Terrett J., Page M.,
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RL	Heparanase Family Member.";
RU	Biochem. Biophys. Res. Commun. 276:1170-1177 (2000).
DR	PIR; JC7506; JCT506.
DR	InterPro: IPR005199; Glyco hydro 79N.
DR	Pfam: PF03662; Glyco hydro 79n; I-
SQ	SEQUENCE 480 AA; 5390 KM; F75F89SF67ACLF83 CRC64;
 Alignment Scores:	
Pred. No.:	3..4e-64 Length: 480
Score:	897.50 Matches: 202
Percent Similarity:	49.20% Conservative: 74
Best local Similarity:	36.01% Mismatches: 146
Query Match:	28.78% Indels: 139
Db:	4 Gaps: 9
 US-10-676-079-3 (1-1721) x Q9HB39 (1-480)	
OY	120 CGCGTGGGTCCTCCCTCCGCGAGGCCTG-----149
DB	18 ProProAlaCysLeuAlaProGlyValLeuTyrlleuAlaLeuLeuHisLeuSerLeu 37
OY	150 -----CCGCAACTGCG-----CAACA 167
DB	38 SerSerGlAlaGlyAAsprArgArgProLeuProValAAsrArgAlaAlaGlyLeuTyrglu 57
OY	168 CAGAGAGTGCTGAGACCTGTGATCTTTCCACCAGAGACCGCTGCACCTGTGAGGCCCTCG 227
DB	58 LysThrLeuLeuLeuLeuAspAlaSerThrLysAsnProValArgThyValAsnGluAsn 77
OY	228 TTCTGTCCGTCAACATTGACCCGAACCTGTGCAGCAGAACCCGCGGTTCTCATCTCTCTG 287
DB	78 PhenSerLeuIleuLeuAspProSerLeileHisAsp---GlyTrpLeuAspPheLeu 96
OY	288 GGTTTCCAAAGCTTGATCCTTGCCCGCAGAGCTTGTTCTCTGGTACCTGAGTTGGT 347
DB	97 SerSerLysArgLeuValThrLeuAlaArgGlyLeuSerProAlaPheLeuArgPheGly 116
OY	348 GGCACCAAGCAGACCTCTTAATTTTT-----GATGCCAAGAAGAATAAAC 395
DB	117 GlyLysArgThrAspPheLeuGlnPheGlnAsnLeuArgAsnProAlaLysSerArg-- 135
OY	396 TTTGAAGAGAAATTACTGCAATCTCAAGTCAACAGATATTGTCMAATATGATCC 455
DB	136 -----GlyGly 137
OY	456 ATCCCTCGTAGTGTGAGAGAAAGTTACGGTTGAATGGCCCTTACAGAGCAATTGCTA 515
DB	138 ProGlyProAsp-----141
OY	516 CTCGAGAAACACTACAGAAAAGTTTCAAGAAGACACTTACAAAGAGCTGTAGAT 575
DB	142 -----TyrtyleuLysAsnTyrglu-----148
OY	576 GTGCTATACACTTTTGCAATGTCTCAGAGACTGACATTGATCTTTGGCTTAATGCGTTA 635
DB	148 -----148
OY	636 TTAAAGAACAGATTTTGAGATGGAACAGTTCTAATGCTAGTTGCTCTGAACTATGC 695
DB	148 -----148
OY	696 TCTTCCAAAGGGTATTAACATTTCTTGGAACTAGGCAATGAACATTCCTTAAAG 755
DB	149 -----AspGluProAsnAsnTyArgThr 156
OY	756 AAGCGTGATATTTTCATCAATGGGTCGAGTTAGAGAAAGATTATATTCATATTCATATA 815
DB	157 Mech1gLyArgAlaValAsnMetSerLeuIleuGlyAspTyrglileuLeuLysSer 176

```

QY 816 CTTCTAAGAAAG---TCCACCTTCAAAAATGCAAACTATGTCCTGATGTTGTCG 872
DB 177 LeuleuGlnProIleArgIleTyrSerKAlaSerLeuTyrGlyProAsnIleGlyArg 196
QY 873 CCTCGAAGAAAGACGGCTTAAGATGCTGAAGACCTTCGAGCGCTGGAGAGATGAT 932
DB 197 ProArgLysAsnValIleAlaLeuLeuAspGlyPheMetLysValAlaGlySerThrVal 216
QY 933 GATTGAGTACATGTCATCTACTACTATTGAATGACCGACTGCTACCGGGAATTTT 992
DB 217 AspaIaValThrTrpGlnHisCysTyrIleAspGlyArgValValLysValMetAspPhe 236
QY 993 CTAACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1052
DB 237 LeuLysThrArgLeuLeuAspThrLeuSerAspGlnIleArgValIleGlnLysValVal 256
QY 1053 GAGACACCGACCGCTGGCAAGAGCTGTGTTAGAGAAACAGCTCTGCATATGAGAGC 1112
DB 257 AsnThrTyrThrProGlyLysLysIleTrpLeuGlnGlyValValThrSerAlaGly 276
QY 1113 GAGAGCGCCCTGCTATCCGACACCTTGGAGCTGTGTTATGTCGTGATTAATTTGGGC 1172
DB 277 GlyThrAsnAsnLeuSerAspSerTyrAlaAlaGlyPheLeuTrpLeuAsnThrLeuGly 296
QY 1173 CTGTGAGCCCGAATGGAATAGAGTGTGATGAGGCAAGTATTCCTTGGAGACGAGAAC 1232
DB 297 MetLeuAlaAsnGlnGlyIleAspValValIleAlaGlnHisSerPhePheAspHisGlyTyr 316
QY 1233 TACCATTTAGTGAATGAAATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1292
DB 317 AsnHisLeuValAspGlnAsnPheAsnProLeuProAspTyrTrpLeuSerLeuLeuTyr 336
QY 1293 AAGAAATTTGTTGGGACCAAGGTGTTATGCGACGCTGCAAGCTTCAAGAGAGAG--- 1349
DB 337 LysArgLeuIleGlyProLysValIleuAlaValHisValAlaGlyLeuGlnArgLysPro 356
QY 1350 -----AAGCTTGGAGTATACCTTCATCTGACCAACCACTGAC 1385
DB 357 ArgProGlyArgValIleArgAspLysLeuArgIleTyrAlaHisCysThrAsnHisHis 376
QY 1386 AATCCAAAGTATTAAGAGAGATTTAACTGTGATGACCAATAACCTCCATTAACCTCAC 1445
DB 377 AsnHisAsnTyrValArgGlySerIleThrLeuPheIleIleAsnLeuHisArgSerArg 396
QY 1446 AAGTACTCCGGTACCTTATCTTTTCTTAACAGCAAGTGGATTAATACCTTTAGAA 1505
DB 397 LysLysIleLysLeuAlaGlyThrLeuArgAspLysLeuValHisGlnTyrLeuLeuGln 416
QY 1506 CTTTGGGACCTCATGATGATTACTTTCATATCTGTCACATCACTGATCTTAACGCTAAG 1565
DB 417 ProTyrGlyGlnGlnGlyLeuLysSerLysSerValGlnLeuAsnGlyLysProLeuVal 436
QY 1566 ATGTGATGATGATCAACCTTGCACCTTATGATGAAACCTCTCCGACGAGAGATTTCA 1625
DB 437 MetValAspArgGlyThrLeuProGlyLeuLysProArgProLeuArgIleGlyArgThr 456
QY 1626 CTGGGCTTCCGACCTTTCATATAGTTTCTTGTGATTAAGAAATGCCAAAGTGTGCT 1685
DB 457 LeuValIleProProValThrMetGlyPhePheValValLysAsnValAsnAlaLeuAla 476
QY 1686 TGC 1688
DB 477 Cys 477

```

```

GN BMEPA.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Bombycidae; Bombyx.
OK NCBI_TaxID=7091;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=p50; TISSUE=Posterior silk gland;
RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
RA Osogawa K., Dejong P.U.;
RT "Genomic sequence of 320kb containing a kettin orthologue on the Z
RT chromosome in Bombyx mori.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RP (2)
RP SEQUENCE FROM N.A.
RC STRAIN=p50; TISSUE=Posterior silk gland;
RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
RA Osogawa K., Dejong P.U.;
RT "Genomic sequence of 320kb containing a kettin orthologue on the Z
RT chromosome in Bombyx mori.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB079860; BAB8191.1; -.
DR EMBL; AB090307; BAC10612.1; -.
DR InterPro; IPR005199; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79n; I.
SQ SEQUENCE 515 AA; 59769 MW; FB8100ABE6DDADB CRC64;

Alignment Scores:
Pred. No.: 8,366-48 Length: 515
Score: 696.00 Matches: 183
Percent Similarity: 50.96% Conservative: 83
Best Local Similarity: 35.06% Mismatches: 182
Query Match: 22.31% Indels: 74
DB: Gaps: 18

US-10-676-079-3 (1-1721) x Q8T108 (1-515)
QY 198 GAGAGCCGCTGACCTGATGAGCCCTGCTTCCTGCTGATGAGCCGACCTG 257
DB 42 GlnGluAspIleLysLeuIleSerGlnAspPheLeuSerPheGlyLysP---ThrIle 60
QY 258 GCCAGGACCCGCGGTTCTTCATCTCTGAGTTCTCCAAAGCTTGTACCTTGGCCAGA 317
DB 61 GlnIleGluAsnTyrAsnArgIleAsnTyrSerAspThrArgLeuArgGluLeuAla 80
QY 318 GCGTGTCTCTGCTGCTGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 377
DB 81 AlaLeuSerProIleArgLeuArgLeuGlyIleThrMetSerGluArgLeuIlePhe--- 99
QY 378 CCCAAGAAAGATCAACCTTTGAGAGAGAGATTACTGCAATCTCAAGTCAACCAAGAT 437
DB 100 ---SerLysGluAsnIle-----ProIleSerCysHis 109
QY 438 ATTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 497
DB 110 AsnLysSerTyrLysSerTyrProLysSerLeuCysGln-----LeuIleGlyLysPro 127
QY 498 TACGAGAGCAATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 557
DB 128 Cys-----LysHisLysHisLysPheLeuProPhePheIleMet 140
QY 558 TCAAGAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 617
DB 141 ThrGlyAsnGlnTyrAsnGlnIleAsnAspPheCysArgLysThrAsnLeuLysLeuLeu 160
QY 618 TTGGCCTAAATGCTTATTAAGACAGACAGATTTGACAGTTCAGTTCATCTGAG 677
DB 161 PheSerLeuAsnAlaMetLeuArgAsp---AsnHisGlyTyrPheAsnGlyLysAsnAlaArg 179
QY 678 TTGCTCTGAGACTGCTCTTCCAGGAGGATTAATCAATTTCTTGGAACTAGGCAATGAA 737
DB 180 GluLeuIleGluPheSerLysHisLysGlnTyrAlaIleAspTrpGlnLeuGlnGln 199

```

RESULT 12  
 Q8T108 PRELIMINARY; PRT; 515 AA.  
 AC Q8T108:  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hepatase-like protein.

```

QY 738 CCGTAAAGTTTCTTAAGAAAGCTGATATTTTCATCAATGGGCTCGCAGTAAAGAGAT 797
Db 200 ProAenSerPheGlnHisValPheAenGluSerValThrProGlnIleLeuAlaLeuAsp 219
QY 798 TAAATTCATATGCAATTAATCTTAAAGAAAGCTTCAAAATGAAATGAAATCTATGCT 857
Db 220 PheGlnIleLeuValGlyValLeuAenHisValGlyValThrGlnSerIleValGly 239
QY 858 CCGTAAAGTTTCTTAAGAAAGCTGATATTTTCATCAATGGGCTCGCAGTAAAGAGAT 908
Db 240 ProAenSerPheGlnHisValPheAenGluSerValThrProGlnIleLeuAlaLeuAsp 259
QY 909 CCGTAAAGTTTCTTAAGAAAGCTGATATTTTCATCAATGGGCTCGCAGTAAAGAGAT 968
Db 260 LeuGlnIleLeuValGlyValLeuAenHisValGlyValThrGlnSerIleValGly 279
QY 969 CCGTAAAGTTTCTTAAGAAAGCTGATATTTTCATCAATGGGCTCGCAGTAAAGAGAT 1028
Db 280 LeuGlnIleLeuValGlyValLeuAenHisValGlyValThrGlnSerIleValGly 296
QY 1029 CCGTAAAGTTTCTTAAGAAAGCTGATATTTTCATCAATGGGCTCGCAGTAAAGAGAT 1082
Db 297 ArgGlnGlnIleGluThrMetGlnAenGlnThrIleValSerIleValSerIleValSer 316
QY 1083 TTAAGAAACCAAGCTTGCATATGAGGCGGAGCGGCTTGTATCGACACCTTGA 1142
Db 317 LeuSerGlnThrSerSerSerIleValGlyValAlaProGlnLeuSerAenThrValAla 336
QY 1143 GCGGCTTATGTCGCGATTAATTTGGGCTGTCACCCGCAATGGGAATGAAAGTGTG 1202
Db 337 GlySerProLeuThrIleLeuValLeuGlyLeuSerIleValSerIleValSerIleVal 356
QY 1203 ATGAGGCAATATTTCTTGAAGCAGAAATACCATTTAGTGAATTAATTCATGCTCT 1262
Db 357 IleArgGlnSerPheIleGly---GlyTyrTyrSerIleValAlaSerIleValSerIle 375
QY 1263 TTAAGTATTTATGCTATCTTCTTGTCAAGAAATTTGGGCGACCAAGTGTATG 1322
Db 376 LeuProAspTrpTrpIleSerValLeuValIleValSerIleValSerIleValSer 394
QY 1323 GCAAGGCTGCAAGTCAAGAAAGAAAGCTTGCATATTCCTTCAACCAAAACT 1382
Db 395 ---GlnAlaGlnCysAenCysSerArgPheGlnArgLeuValIleHisCysThrAenArg 413
QY 1383 GACAATTCAGATTAAGAA---GGAATTTAATCTGTATGCGCAATAAC--- 1430
Db 414 ---LysTyrThrAenAspThrSerAlaValIleValSerIleValSerIleValSer 430
QY 1431 ---CTCATTAAC---GTC 1442
Db 431 MetAlaIleValAlaArgPhePheLeuAenGlyThrAlaLeuHisGlyAerAspLeuIle 450
QY 1443 ACCAAGTATTCGCGTAACTCTTCTTCTTAAGAAAGAAAGTGAATTAATCTCTTA 1502
Db 451 HisGlnIleValIle---IleSerAlaProSerAenHisValSerIleValSerIleVal 463
QY 1503 AGACCTTTGGAGCTCATGATTAATTTCAAAATCTGTCAATCTCATGCTTAATCTTA 1562
Db 464 ---SerLysThrIleLeuLeuAenGlyTrpProLeu 474
QY 1563 AAGATGATGATATCAAACTTGCACCTTTAAATGAAAAAAGCTTCGCGGCGAGAAAT 1622
Db 475 TyrTyr---GluSerAenLeuHisValLeuValArgProAenIleHisArgTyrGlyArg 492
QY 1623 TCACTGGGCTTCGACGCTTCTCATATAGTTTTTTTGTGATTAAGAAATGCGAAATGCT 1682
Db 493 TyrValSerLeuProTyrSerIleGlyPheThrValIleValSerIleThr 512
QY 1683 GCTTGC 1688
Db 513 ValCys 514

```

```

RESULT 13
Q9SDA1 PRELIMINARY; PRT; 521 AA.
AC Q9SDA1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN F13G24.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RX [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RA SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: A133421; CAB62595.1; -.
DR PIR: T45608; T45608.
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR005199; Glyco hydro 79N.
DR InterPro: IPR001254; Peptidase S1.
DR Pfam: PF03662; Glyco hydro 79n; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Alignment Scores:
Pred. No.: 4,85e-25 Length: 521
Score: 416.00 Matches: 154
Percent Similarity: 42.05% Conservative: 68
Best Local Similarity: 29.17% Mismatches: 184
Query Match: 13.34% Indels: 122
DB: 10 Gaps: 24

US-10-676-079-3 (1-1721) x Q9SDA1 (1-521)
QY 285 CCGGTTCTTCCAAAGCTTGTACCTTGGCCAGAGGCTTCTCTGTAAGCTT 344
Db 55 LeuThrArgProLeuValIleValAlaIleValAlaPheValSerLeuValGlyIle 74
QY 345 GGTGGCAACCAAGCACTTCTTAATTTTCGATCCCAAGAAATCAACCTTGAAG 404
Db 75 GlyGlySerLeuGlnAspGlnValIleTyrAspValGlyValLeuValSerIle 91
QY 405 AGAAGTTACTCGCAATCTCAAGTCAACAGATTTGAAATATGATTCATCTCTCT 464
Db 92 ---ProCysArg--- 94
QY 465 GATGTGAGAGAAAGTTAAGGTTGGAATGCGCTTACAGAGCAATTTGCTACTCCAGA 524
Db 95 ---ProPheGlnLysMet--- 99
QY 525 CACTACCAAAAAAGTTCAAGAAACAGC---ACCTACTCAAGAAAGCTCTGA--- 572
Db 100 ---AenSerGlyLeuPheGlyPheSerIleValSerIleValSerIleVal 112
QY 573 ---GATGTGATATCACTTTTTCGAAATCTGCGAGGACTGATCTT 620
Db 113 MetLysArgTrpAspGlnLeuAenSerPheLeuThrAlaThrGlyAlaValThrPhe 132
QY 621 GGCCTAAATGCGTTATTAAGAACAGACAGATTTGAC---TGG 659
Db 133 GlyLeuAenAlaLeuValGlyValArgHisValSerValAlaThrGlyValAlaTrp 152
QY 660 AACAGTTCTAATGCTCAAGTTGCTCTGACCTACTGCTTCCAAAGGATTAACATT--- 716

```

```

Db      153 Asph1s1leantThrGlnasphPheleuentyrThrValserIyGlyTyrrValIleasp 172
Qy      717 TTTTGGAACTGAGGCAAGCACTTAACAGCTTCTTAAGAGGCTGATATTTTCATCAT 776
Db      173 SerTrpGluPheGlyAsnGluLeuSerGly-----SerGlyValGlyAlaSerValSer 190
Qy      777 GGGTCGACGATTAGAGAGATTATATTCATTAACCTTAAGAAAGCCACCTTC 836
Db      191 AlaGluLeuTyrglyLeuAspLeuIleValLeuIleAspValIleAsnIle---ValIyr 209
Qy      837 AAAAATGCAGAACTTATGCTTCGATGCTGATGCTGACCTT-----CGAAGA 881
Db      210 LysAsnSerTrpLeuIleIySerProIleLeuValAlaProGlyGlyPheTyrgluGln 229
Qy      882 AAGACGGCTAAGATGCTGAAGAGCTTCTCGAAGCGCTGGAGGAAGTATTCATTCAGT 941
Db      230 TrpTyrrTrpIyLeuLeuGluIle-----SerGlyProSerValValAlaSerValVal 246
Qy      942 ACATGGCATCTACTATTTGAATGAGCGAGCT-----GCTACGAGGAGATTTTCTA 995
Db      247 ThrHisIleTyrrAsnLeuGlySerGlyAsnAspProAlaLeuValIyIleIleIec 266
Qy      996 AACCTGATGATATGAGACATTTTATTCATCTGTCGAAAGATTTTC-----1043
Db      267 AspProSer-----TyrLeuSerGlnValSerIyThrPheIyAspValAsn 282
Qy      1044 CAGGTGGTTGAGAGACACCGGCTGGCAAGAGCTGTGTTGAGGAAGAACCTCTGCA 1103
Db      283 GlnThrIleGlnGluIleGlyProTrpAlaSerProIyrValGlyIleSerGlyAla 302
Qy      1104 TATGAGAGCGAGCGCCCTTGATCTCGACCTTTCGACGCTGAGCTTATGNGCGTGA 1163
Db      303 TyrAsnSerGlyGlyArgHisValSerAspTrpPheIleAspSerPheTrpTyrrLeuAsp 322
Qy      1164 AAATTTGGGCTTCAGCCCGAATGGGAATGAGTGTGATAGGCAAGATATTTCTTTGA 1223
Db      323 GlnLeuGlyMetSerAlaArgHisAsnThrIyValTyrrCysArgIleThrIleValGly 342
Qy      1224 GCAGAAACTACCATTTAGTGATGAA---AACTTCGATCCTTTACCTGATATTTGGCTA 1280
Db      343 ---GlyPheTyrglyLeuLeuGluIySerGlyThrPheValProAspProAspTyrrIySer 361
Qy      1281 TCTCTTGTGTGAAGAAATTTGGGGGACCAAGGTGTTATATGGAAGCGTGAAGTTCA 1340
Db      362 AlaLeuLeuTrpHisArgLeuMetGlyIyGlyValLeuAlaValGlnThrAspGlyPro 381
Qy      1341 AAGAGAAAGAGACCTTCAGATATACCTTCATTCACAAACATGACATCCAAAGTATAA 1400
Db      382 Pro-----GlnLeuArgValTyrrAlaHisCysSerIyS-----Gly 393
Qy      1401 GAAGAGATTTAATCTGTATGCCATTAACCTTCATACGTCACCAAGTACTTG-----1454
Db      394 ArgAlaGlyValThrLeuLeuLeuIleAsnLeuSerAsnGlnSerAspPheThrValSer 413
Qy      1454 -----1454
Db      414 ValSerAsnGlyIleAsnValValLeuAsnAlaGluSerArgIyLeuIyIySerLeuLeu 433
Qy      1455 ---CGTTACCTTATCTTTTCT-----AACAGCAAGTGGATTAATACCTTTCTA 1502
Db      434 AspThrIleLeuIyAspProPheSerTrpIleGlySerIyAlaSerAspGlyTyrrIleAsn 453
Qy      1503 AAGACT-----TTGGACCTCATGGA-----TTCCTTCCAAATCTGTCCACTC 1547
Db      454 ArgGluGluTyrrHisIleLeuThrProGluIleAsnGlyValLeuArgSerIyThrMetValLeu 473
Qy      1548 AATGCTAATCTTAAGATGGATGATGATCAACCTTGCAACCTTATATGAGAAACCT 1607
Db      474 AsnGlyIySerLeuIySerProThrAlaThrGlyAlaAspIleProSerIeu---GluProVal 492
Qy      1608 CTCGCGCCA---GGAATTCACCTGGGCTTGCACAGCTTCTCATATATGTTTTTGTGATA 1664

```

```

Db      493 LeuArgSerValAsnSerProLeuAsnValLeuProLeuSerMetSerPheIleValLeu 512
Qy      1665 AGAATGCCAAGTTCGTCGCTTGC 1688
Db      513 ProAsnPheAspAlaSerAlaCys 520
RESULT 14
ID      09PF10      PRELIMINARY;      PRT;      543 AA.
AC      09PF10;
DT      01-MAR-2001 (TREMBLrel. 16, Created)
DT      01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Similarity to heparanase.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=Columbia;
RX      MEDLINE=97471969; PubMed=9330910;
RA      Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA      Miyajima N., Tabata S.;
RT      "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT      features of the 1.6 Mb regions covered by twenty physically assigned
RT      pl clones."
RL      DNA Res. 4:215-230 (1997).
DR      EMBL; AB005249; BAB09947.1;
DR      GO; GO:0004295; F:trypsin activity; IEA.
DR      GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR      InterPro; IPR005199; Glyco_hydro_79N.
DR      InterPro; IPR001254; Peptidase_S1.
DR      Pfam; PF03662; Glyco_hydro_79n; 1.
DR      PROSITE; PS00135; TRYPSIN_SER; 1.
SQ      SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;

Alignment Scores:
Pred. No.:      4.92e-25      Length:      543
Score:      416.00      Matches:      154
Percent Similarity:      42.05%      Conservative:      68
Best Local Similarity:      29.17%      Mismatches:      184
Query Match:      13.34%      Indels:      122
DB:      Gaps:      24

US-10-676-079-3 (1-1721) x 09PF10 (1-543)
Qy      285 CTGGGTTCTCCAAAGCTTCGTACCTTGGCGAGAGGCTGTCTCGTACCTGAGGTTT 344
Db      77 LeuThrArgProLeuLeuThrIyAlaIleIyAlaIleIyAlaPheIySerProLeuArgIleArgIle 96
Qy      345 GGTGGACCAAGACAGACTTCTTAATTTTTCATCCCAAGAAAGATCAACCTTTGAAGAG 404
Db      97 GlyIySerLeuGlnAspGlnValIleTyrrAspValGlyAsnLeuIyThr-----113
Qy      405 AGAAGTTACTGGCAATCTCAAGTCAACCAAGATATTTGCAATATGATTCATCCCTCTCT 464
Db      114 -----ProCysArg-----116
Qy      465 GATGTGAGAGAGAGATTACGGTTGGAATGGCCCTTACAGAGCAATGCTACTCCGAGAA 524
Db      117 -----ProPheGlnIyMet-----121
Qy      525 CACTACAGAAAAAGTTCAAGACAGC-----ACCTACTCAGAGAGCTTGTA---572
Db      122 -----AsnSerGlyLeuPheGlyPheSerIyGlyCysLeuHis 134
Qy      573 -----GATGCTATACACTTTTGCAAACGCTCAGAGACTGATATCTT 620
Db      135 MetIySarGlyTrpAspGluLeuAsnSerPheLeuThrAlaThrGlyAlaValIyThrPhe 154
Qy      621 GGCCTTAATGCGTTATTAAGAACAGACAGATTGCGAG-----TGG 659

```



```

Db      155 GlyLeuAsnAlaLeuArgGlyArgHsLysLeuArgGlyLysValatrpGlyValatrp 174
Qy      660 AAGAGTTCTAATGCTGAGTTCCTGAGCTACCTGCTTCCAAAGGGGTTAAACATT--- 726
Db      175 AspHisIleAsnThrGlnAspPheLeuAsnThrValSerIleGlyValIleAsp 194
Qy      717 TCTGGAACTAGCAATGAACCTAACAGTTTCCTTAAGAGGCTGATATTTTCATCAAT 776
Db      195 SerTrpGluPheGluAsnGluLeuSerGly-----SerGlyValGlyAlaSerValSer 212
Qy      777 GGGTCGAGTTAGAGAAAGATTAATATTCATTCGATTAACCTTCAAGAAAGTCCACTTC 836
Db      213 AlaGluLeuTyrGlyLysAspLeuIleValLeuLysAspValIleAsnLys---ValTyr 231
Qy      837 AAAAATGCAAAATCTATGCTGCTGATGCTGATGCTGACGCT-----CGAAGA 881
Db      232 LysAsnSerTrpLeuIleLysProIleLeuValAlaProGlyGlyPheTyrGluGlnGln 251
Qy      882 AAGACGGCTAAGATGCTGAAGACCTCTGAAGGCTGGAGAGAGTGAATGATTCAGTT 941
Db      252 TrpTyrThrLysLeuLeuGluIle-----SerGlyProSerValIleAspValVal 268
Qy      942 ACATGGATGCACTATTTTGAATGACGAGCT-----GCTACCGAGAAAGATTTTCTA 995
Db      269 ThrHisIleIleTyrAsnLeuGlySerGlyAsnAspProAlaLeuValLysLysIleMet 288
Qy      996 AACCTGATGATTAATGACATTTTATTTTATTTCTGTCGAAAAGTTTTC----- 1043
Db      289 AspProSer-----TyrLeuSerGlnValSerLysThrPheLysAspValAlaAsn 304
Qy      1044 CAGGTGTTAGAGCACCAGGCTGGCGAAGAGGCTGTTAGAGAAACAGCTGCA 1103
Db      305 GlnThrIleGlnGluIleGlyProTyrAlaSerProTyrValGlyGluSerGlyValAla 324
Qy      1104 TATGAGGCGGAGGCGCTGCTATCCGACACCTTTCAGCTGCTTATGAGGCTGAT 1163
Db      325 TyrAsnSerGlyLysArgHsIleValSerAspThrPheIleAspSerPheTyrLysAsp 344
Qy      1164 AAATGGGCTGTCAGCCGCAATGGGAATGAGTGTGATGAGGCAAGTATCTTTGGA 1223
Db      345 GlnLeuGlyMetSerIleAlaArgHsIleAsnThrLysValTyrCysArgGlnThrLeuValGly 364
Qy      1224 GCAAGAACTACCACTTATGAGTGAAG---AAGCTGATGCTTATCTGATTAATGGCTA 1280
Db      365 ---GlyPheTyrGlyLeuLeuGluLysGlyThrPheValIleProAsnProAspTyrTyrSer 383
Qy      1281 TCTCTTGTATCAAGAAATGGTGGGCAACAAGTGTATTAAGCAACGTCGAAGGTTCA 1340
Db      384 AlaLeuLeuTrpHisArgLeuMetGlyLysGlyValAlaLeuAlaValGlnThrAspGlyPro 403
Qy      1341 AAGAGAGAGAGCTTCGATGATACCTTATTCGACAAACACTGACAAATCCAAAGTATMAA 1400
Db      404 Pro-----GlnLeuArgValTyrAlaHisCysSerLys-----Gly 415
Qy      1401 GAAAGAGATTAATCTGTATGCCATTAACCTCCATTAAGCTACCAAGTACTTG----- 1454
Db      416 ArgAlaGlyValThrLeuLeuLeuIleAsnLeuSerAsnGlnSerAspPheThrValSer 435
Qy      1454 ----- 1454
Db      436 ValSerAsnGlyLysAsnValValLeuAsnAlaGluSerArgLysLysLysSerLeuLeu 455
Qy      1455 ---CGGTACCTATCTTTTCT-----AACAAAGCAAGTGTAAATATCTTCTA 1502
Db      456 AspThrLeuLysArgProPheSerTrpIleGlySerIleValAspArgLysTyrLeuAsn 475
Qy      1503 AGACCT-----TTGGACCTCATGAG-----TTACTTCCAAATGTGCCAACTC 1547
Db      476 ArgGluGluTyrHisLeuLeuTrpProGluAsnGlyValLeuArgSerLysThrMetValLeu 495
Qy      1548 AATGGTCTAATCTCAAGAGTGGATGATGACCAACCTTGCCACCTTAAAGGAAACCT 1607

```

```

Db      496 AsnGlyLysSerLeuLysProThrAlaThrGlyAspIleProSerLeu---GluProVal 514
Qy      1608 CTCGGGCCA---GGAAGTTCACTGAGGCTGGCCAGCTTCTCATATAGTTTTTTTGATA 1664
Db      515 LeuArgSerValAsnSerProLeuAsnValLeuProLeuSerMetSerPheIleValLeu 534
Qy      1665 AGAAATGCCAAAGTTCCTGCTTGC 1688
Db      535 ProAsnPheAspAlaSerAlaCys 542

RESULT 15
O89F99
ID O89F99 PRELIMINARY; PRT; 559 AA.
AC O89F99;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DE 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE BR6802. protein.
GN BR6802.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneo T., Nakamura Y., Sato S., Minamiyawa K., Uchiumi T.,
RA Sasamoto S., Matsumura A., Idegawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Teurloka H., Wada T., Yamada M.,
RA Tsubata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005959; BAC52067.1; -.
KW Complete proteome.
SQ SEQUENCE 559 AA; 59948 MW; F3B423547AA30170 CRC64;

Alignment Scores:
Pred. No.: 2,94e-24 Length: 559
Score: 406.50 Matches: 164
Percent Similarity: 41.57% Conservative: 80
Best Local Similarity: 27.94% Mismatches: 211
Query Match: 13.03% Indels: 132
DB: 16 Gaps: 28

US-10-676-079-3 (1-1721) x O89F99 (1-559)
Qy      62 GATGCTGCTGCGTCCGAAGCTGCGTCCGCGCGCTGATGCTGCTGCTGCGGCC 121
Db      60 AspAlaIleArgValThrGlyLeuAlaLeuLeuAlaIleGlyThrSerCysAlaIleHis 79
Qy      122 GCTGGG-----TCCCTGCTCCCTGCGCGCTGCGCCGCGGACCTGCGCAGACAGAGCGT 175
Db      80 AlaGlyThrValSer-IleAlaProAlaGlyLeu----- 90
Qy      176 CGTGAAGCTGACTTCTTCAACCGAGCGCGCTGCAACCTGGTGAAGCCCTGCTCTGTC 235
Db      91 -----ArgAlaIleGlyThrIleAspProArgPheGlnSe 102
Qy      236 CCGTACCATTTGAGCGCAACCTGCGCAGACCGCGCGGTTCCTC----- 278
Db      102 TTYrAsnIleGluMetValGluValThrGlyArgPheTrpLysProTyrProGlnAl 122
Qy      279 -----ATCCTCGTGGGCTC 292
Db      122 aMetArgAlaTrpAlaAspLysAspArgTyrSerTyrArgProProIleAspLeuGlyAs 142
Qy      293 TCCAAAGCTTGTACCTTGGCCAGAGCTTGTCTCTCGGCGTACCTGAGGTTGGTGCGAC 352
Db      142 nThrArgLeuArgGlyLeuAlaValAlaLeuSerProAlaTyrLeuArgValSerGlyTh 162
Qy      353 CAAGACAGACTTCTTAATTTTGCATCCCAAGAAAGATCAACCTTTGAAGAGAAAGTTA 412

```

```

Db      466  B-----HisProser---LysArgGlyAlaValThrValLeuAlaIleAsnI 481
Qy      1433 CCATTAAGCTCAC--AAGTACTTGGCGGTACCTTATCTTTTCTTACAAGCAAGTGA 1489
          |||      |||      |||      |||      |||
Db      481  eserArgSerThrIalArgThrIleValLeuProLeuP--AlaGl 496
          |||      |||      |||      |||      |||
Qy      1490 TAAATTAACCTTCTAGACCTTTGGGACCTCATGATTACTTCCAAATCTGTCGAATCCA 1549
          |||      |||      |||      |||      |||
Db      496  uArgTylThrIeuInIalAlaArgLeuGInGly-----AlaThrValGInLeuAs 513
          |||      |||      |||      |||      |||
Qy      1550 TGGCTTAACCTTAAGAAGTGTGATGATCAAACTTGCACCTTTAATGGAAGAAACCTCT 1609
          |||      |||      |||      |||      |||
Db      513  nGlyIyThrIeuAlaIeuThrIaArgGluIaArgGlyLeuProProIeuAlaGlyAlaI 533
          |||      |||      |||      |||      |||
Qy      1610 CCGGCAAGAAAGTTCACTGCGCTTCCAGACTTCTCATATAGTTTTTGTGATTAAGAA 1669
          |||      |||      |||      |||      |||
Db      533  eGluAlaGly---AlaValGInLeuAlaSerGluThrIleThrPheLeuAlaIleProGl 552
          |||      |||      |||      |||      |||
Qy      1670 TGCCAAAAGTCTGCTTGC 1688
          |||      |||      |||      |||
Db      552  yAlaAlaAsnSerAlaCys 558
          |||      |||      |||      |||

```

Search completed: August 28, 2004, 07:51:40  
Job time : 248.5 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 28, 2004, 07:42:36 ; Search time 193 Seconds  
(without alignments)  
5039.008 Million cell updates/sec

Title: US-10-676-079-3  
Perfect score: 3119  
Sequence: 1 ctgagcttcgactctccg.....atactagctcctgactc 1721

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 28547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

MODEL=frame\_n2p.model -DEV-rlp  
-O=/cgn2.1/USPTO\_spool\_p/US10676079/runat\_28082004\_084224\_23795/app\_query.fasta.1.1863  
-DB=A\_Geneseq\_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10676079 @CGN 1.1.263 @runat\_28082004\_084224\_23795 -NCRU=6 -ICPU=3  
-NO MMAP -LANG=QUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

A\_Geneseq\_29Jan04:\*  
1: geneeqp1980s:\*  
2: geneeqp1990s:\*  
3: geneeqp2000s:\*  
4: geneeqp2001s:\*  
5: geneeqp2002s:\*  
6: geneeqp2003as:\*  
7: geneeqp2003bs:\*  
8: geneeqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2949	94.5	592	2	AA02346 A human h
2	2949	94.5	592	3	AA08850 Amino aci
3	2904	93.1	588	2	AA030124 A human p
4	2842	91.1	543	2	AA02345 A human h
5	2842	91.1	543	3	AA05290 Human hep
6	2842	91.1	543	3	AA08849 Amino aci
7	2842	91.1	543	3	AA05290 Human hep
8	2842	91.1	543	4	AA07635 Human hep
9	2842	91.1	543	5	AB07813 Human hep
10	2838	91.0	543	2	AA017082 Human hep

11	2838	91.0	543	4	AA086206 Human hep
12	2838	91.0	543	7	ADD18950 Human dis
13	2826	90.6	543	4	AA088361 Human mem
14	2817	90.3	545	6	ABP56822 Human hep
15	2817	90.3	545	7	ABE16012 G-coupled
16	2772	88.9	532	2	AA017083 Seg ID No
17	2764	88.6	530	2	AA034173 Human pre
18	2673.5	85.7	527	5	AB007815 Chicken s
19	2146	68.8	535	3	AA08851 A murine
20	2146	68.8	535	5	AB007811 Mouse hep
21	2123	68.1	536	5	AB007812 Rat hep
22	1645.5	52.8	523	3	AB007814 Chicken h
23	1614	51.7	530	2	AA017085 Rat hep
24	1602	51.4	380	2	AA017084 Mouse hep
25	1154.5	37.0	592	4	AA07632 Human hep
26	1154.5	37.0	592	4	AA07632 Human hep
27	1148.5	36.8	592	4	AA081062 Human hep
28	1147.5	36.8	592	4	AA085215 Hepatarnas
29	1142.5	36.6	582	5	AAE18326 Human hep
30	1112.5	35.7	538	4	AA097633 Human hep
31	1106.5	35.5	528	5	AAE18327 Human hep
32	936.5	30.0	534	4	AA085216 Hepatarnas
33	936.5	30.0	534	5	ABP69310 Human pol
34	936.5	30.0	534	5	AA050337 Human pre
35	927.5	29.7	492	4	AA084664 Amino aci
36	897.5	28.8	480	4	AA097634 Novel hum
37	897.5	28.8	480	4	AA07634 Human hep
38	897.5	28.8	480	4	AA085217 Hepatarnas
39	892.5	28.6	470	5	AAE18328 Human hep
40	891.5	28.6	439	4	AA007423 Human hep
41	788	25.3	331	5	AA050383 Human hep
42	663	21.3	488	4	AA031469 Amino aci
43	645	20.7	488	4	AA031470 Amino aci
44	642	20.6	488	4	AA031472 Amino aci
45	622	19.9	488	4	AA031471 Amino aci

## ALIGNMENTS

RESULT 1	AA02346	standard; protein; 592 AA.
ID	AA02346	
AC	AA02346	
XX		
DT	09-JUL-1999	(first entry)
XX		
DE	A human heparanase protein.	
XX		
KW	Heparanase; hpa; modulator; heparin-binding growth factor;	
KW	cellular response; cytokine; cell interaction; plasma lipoprotein;	
KW	cellular susceptibility; infection; disintegration;	
KW	neurodegenerative plaque; wound healing; angiogenesis; restenosis;	
KW	atherosclerosis; inflammation; neurodegenerative disease; neuritis;	
KW	plasma heparin; micrometastasis; autoimmune lesion; renal failure.	
XX		
OS	Homo sapiens.	
XX		
PN	W09911798-A1.	
XX		
PD	11-MAR-1999.	
XX		
PF	31-AUG-1998; 98MO-US017954.	
XX		
PR	02-SEP-1997; 97US-00922170.	
XX		
PR	02-JUL-1998; 98US-00109386.	
XX		
PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.	
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.	
PA	(FRIE/) FRIEDMAN M M.	
XX		
PI	Pecker I, Vlodavsky I, Feinstein E,	
XX		

DR WPI; 1999-302255/25.  
DR N-PSDB; AAX35650.

XX New human polynucleotide useful for treating angiogenesis, restenosis,  
PT and inflammation.  
XX

PS Claim 6; Page 65-66; 63pp; English.

XX The specification describes a polypeptide having heparanase (hpa)  
CC activity. The recombinant protein is used as a modulator of heparin-  
CC binding growth factors, cellular responses to heparin-binding growth  
CC factors and cytokines, cell interaction with plasma lipoproteins,  
CC cellular susceptibility to viral, protozoal and bacterial infections or  
CC disinfection of neurodegenerative plaques. Heparanase may be useful for  
CC conditions such as wound healing, angiogenesis, restenosis,  
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral  
CC infections. Mammalian heparanase can be used to neutralize plasma  
CC heparin, and anti-heparanase antibodies may be applied for  
CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and  
CC renal failure in biopsy specimens, plasma samples, and body fluids. The  
CC present sequence represents human heparanase

SO Sequence 592 AA;

Alignment Scores:  
Pred. No.: 2,98e-283 Length: 592  
Score: 2949.00 Matches: 562  
Percent Similarity: 99.82% Conservative: 0  
Best Local Similarity: 99.82% Mismatches: 1  
Query Match: 94.55% Indels: 0  
DB: 2 Gaps: 0

US-10-676-079-3 (1-1721) x AAY02346 (1-592)

QY 3 AGAGCTTTCAGCTCTCGCGCGGAGCTGCGGGGGAGCAGCCAGGTGAGCCCAAG 62  
DB 30 ArgAlaLeuAspSerProLeuArgGlySerTrpArgIleGlnProGlyGlnProIle 49  
QY 63 ATGCTGCTGCTCGAGCCTGCGCTGCGCGCGCGCTGATGCTGCTCTCGGCGCG 122  
DB 50 MetLeuLeuArgSerIleProAlaLeuProProIleMetLeuLeuGlyPro 69  
QY 123 CTGGGTCCT 182  
DB 70 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAlaValAsp 89  
QY 183 CTGAGCTTCTTCAACCCAGAGCGCGCTGACCTGTGAGCCCTGCTCTCTCTCTCT 242  
DB 90 LeuAspPhePheThrGlnGlnProLeuHleuValSerProSerPheLeuSerValThr 109  
QY 243 ATTGACGCCAAGCTGCGCAGGAGCCCGGTTCTCTCTCTCTCTCTCTCTCTCTCT 302  
DB 110 IleAspAlaLeuLeuAlaThrAspProArgPheLeuIleLeuGlySerProIleLeu 129  
QY 303 CGTACCTTGGCCAGAGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 362  
DB 130 ArgThrLeuAlaArgGlyLeuSerProAlaIleuArgPheGlyIleThrIleAsp 149  
QY 363 TTTCTTATTTTTCGATCCCAAGAGGATCAACCTTGAAGAGAGATTACTGGCATCT 422  
DB 150 PheLeuIlePheAspProIleGlySerIleThrPheGlnGlnIleGlnSerIle 169  
QY 423 CAAGTCAACAGATATTTTGAATATGATTCATCTCTCTCTCTCTCTCTCTCTCTCT 482  
DB 170 GlnValAsnGlnAspIleCysIleTyrcIleSerIleProAspValGlnGlnIleLeu 189  
QY 483 CGGTGGATAGCGCTTCAACAGAGCAATGCTACTCCGAGAACTCAACGAAAAAGTTC 542  
DB 190 ArgLeuGlnIleTrpProIleGlnGlnIleLeuLeuArgGlnIleIleTyrcIleValPhe 209  
QY 543 AAGAAAGCAAGCTCTGATGAAGCTGTGATGATGATGATGATGATGATGATGATGAT 602  
DB 210 LysAsnSerThrIleArgSerSerValAspValLeuIleThrPheAlaAsnCysSer 229

QY 603 GCACTGAGCTGATCTTTGGCCCTAAATGCGTTATTAAGAACAGCAGATTGTCAGTGAAC 662  
DB 230 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn 249  
QY 663 AGTTCTTAATGCTCAGTTGCTCTGAGCTACTGCTCTCCAGAGGGTATTAACATTTCTTGG 722  
DB 250 SerSerAsnAlaGlnLeuLeuLeuAspIleCysSerSerIleGlyIleAsnIleSerTrp 269  
QY 723 GAATGAGCAATGATCACTTAACAGTTCTTAAAGAGCGTGAATTTTCAATGAGGTG 782  
DB 270 GlnLeuGlyAsnGlnProAsnSerPheLeuIleValAspIlePheIleAsnGlySer 289  
QY 783 CAGTTAGAGAGATTAATTTCAATTGCAATTAACCTTAAAGAGTCCACTTCAAAAT 842  
DB 290 GlnLeuGlyGlnAspIleGlnIleAsnIleValAspLeuLeuArgIleSerThrPheIleAsn 309  
QY 843 GCMAAACTATATGCTCTGATGTTGGTCAAGCTCCAGAAAGAGCGCTTAAGTCTGAAG 902  
DB 310 AlaIleLeuTyrcIleProAspValGlnGlnProArgIleThrAlaIleMetLeuIle 329  
QY 903 AGCTTCTGAGGCTGCTGAGAGAGATGATGATTCAGTTACATGAGCTACTATTTG 962  
DB 330 SerPheLeuIleValGlyIleGlnValIleAspSerValThrTrpIleIleIleValLeu 349  
QY 963 AATGACGAGCTGCTACACAGAGAGATTTTCTAAACCTGATGATGACATTTTAT 1022  
DB 350 AsnGlyArgThrAlaThrArgIleGlnAspPheLeuAsnProAspValLeuAspIlePheIle 369  
QY 1023 TCATCTGTGMAAAAGTTTCCAGGTGTTGAGAGACACAGGCTGTGCAAGAGCTGTG 1082  
DB 370 SerSerValGlnIleValPheGlnValIleGlnSerThrArgProIleIleValIleTrp 389  
QY 1083 TTAGAGAAACAAAGCTCTGATATGAGAGCGAGCGCCCTGCTATCCGACACTTTGA 1142  
DB 390 LeuGlyGlnThrSerSerIleAlaTyrcIleGlyIleAlaProLeuLeuSerAspThrPheAla 409  
QY 1143 GCTGGCTTATATGCTGATTAATTTGGGCTGTGACCCGAAATGGGAAATGAAGTGTG 1202  
DB 410 AlaGlyPheMetTrpLeuAspIleGlyLeuSerAlaArgIleCysIleGlnValIle 429  
QY 1203 ATAGGCAAGTATCTTTTGAAGAGAGAACTACATTTAGTGAATGAATTTCCATCTCT 1262  
DB 430 MetArgGlnValPhePheGlyAlaGlyAsnIleTyrcIleValAlaAspIleAsnPro 449  
QY 1263 TTACTGATTAATGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1322  
DB 450 LeuProAspTrpTrpLeuSerLeuLeuPheIleValGlyIleThrIleValLeuMet 469  
QY 1323 GCAAGCGTGCMAAGTTCAAGAGAGAAAGCTTGAAGTATACCTTATGCAAAACCT 1382  
DB 470 AlaSerValGlnIleGlySerIleAspArgGlyLeuArgValIleTyrcIleIleCysThrAsnThr 489  
QY 1383 GACATTCMAAGTATTAAGAGAGATTTTAATCTGTATAGCCATAAACCCTCAATACGTC 1442  
DB 490 AspAsnProArgIleTyrcIleGlnIleAspLeuThrIleTyrcIleAlaIleAsnIleAsnVal 509  
QY 1443 ACCAAGTACTTGGGTTACCTTACTCTTTTCTTAACAACAAGTGAATTAATCTCTCT 1502  
DB 510 ThrIleTyrcIleLeuArgLeuProIleTrpProPheSerAsnIleGlnValAspIleTyrcIleLeu 529  
QY 1503 AGACTTTGGAGCTCATGATTAATCTTCAAAATCTGTCCAATCAATGATCTTAATCTCT 1562  
DB 530 ArgProLeuGlyProIleGlyLeuLeuSerIleSerValGlnLeuAsnGlyLeuThrLeu 549  
QY 1563 AAGATGTGATGATCAAACTTTGCCACTTTTAATGAGAAAACTTCCGCGCAGAGAGT 1622  
DB 550 LysMetValAspAspGlnThrLeuProProLeuMetGlyIleProLeuArgProGlySer 569  
QY 1623 TCATGAGGCTGCTGACCTTCTCATATAGTTTCTTGTATGAAGAAATGCAAGTGTGCT 1682  
DB 570 SerLeuGlyLeuProAlaPheSerIleSerPhePheValIleArgAsnAlaIleValAla 589

QY 1683 GCTTCATC 1691  
 DB 590 AlaGville 592

RESULT 2  
 AAB08850  
 ID AAB08850 standard; protein; 592 AA.  
 AC AAB08850;  
 XX 15-JAN-2001 (first entry)  
 DT 15-JAN-2001 (first entry)

XX Amino acid sequence of a human heparanase polypeptide.  
 XX Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;  
 XX heparin-binding growth factor; cytokine; neurodegenerative plaque;  
 XX wound healing; infection; burn; angiogenesis; restenosis;  
 XX atherosclerosis; inflammation; neurodegenerative disease;  
 XX Gerstmann-Strausler Syndrome; Creutzfeldt-Jakob disease.  
 XX Homo sapiens.  
 XX WO200052178-A1.  
 XX 08-SEP-2000.  
 XX 14-FEB-2000; 2000WO-US003542.  
 XX 01-MAR-1999; 99US-00258892.  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX (FRIE/) FRIEDMAN M M.  
 XX Pecker I, Vlodayevsky I, Feinstein E;  
 DR WPI; 2000-579289/54.  
 DR N-PSDB; AAA75053.

XX New polynucleotides encoding a polypeptide having heparanase activity,  
 PT useful in wound healing and in gene therapy, particularly in treating  
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.  
 PS Claim 22; Page 122-123; 152pp; English.

XX The present sequence represents a human protein with heparanase catalytic  
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,  
 CC particularly in treating tumor, inflammation or autoimmunity.  
 CC Particularly, the polynucleotide is useful in modulating the  
 CC bioavailability of heparin-binding growth factors, cellular responses to  
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.  
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular  
 CC susceptibility to certain viral and some bacterial and protozoa  
 CC infections, or disintegration of neurodegenerative plaques. The  
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or  
 CC radiation burns), and in the treatment of angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-  
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,  
 CC bacterial or protozoa infections

XX Sequence 592 AA;  
 SQ

Alignment Scores:  
 Pred. No.: 2,98e-283 Length: 592  
 Score: 2949.00 Matches: 562  
 Percent Similarity: 99.82% Conservative: 0  
 Best Local Similarity: 99.82% Mismatches: 1  
 Query Match: 94.55% Indels: 0  
 DB: 3 Gaps: 0

US-10-676-079-3 (1-1721) x AAB08850 (1-592)

QY 3 AGAGTTTGCAGCTCCGCTGCGCGGCACTGCGCGGGGAGACAGCCAGGTAGCCCAAG 62

DB 30 ArgAlaLeuAapSerProLeuArgGlySerTrpArgGlyGluGlnProGlyGluProIys 49  
 QY 63 ATGCTGCTGCGCGCTGCAACCTGCGCGCGCGCGCGCTGATGCTGCTCTCTGGGGCG 122  
 DB 50 MetLeuLeuAArgSerLysProAlaLeuProProProLeuMetLeuLeuLeuLysPro 69  
 QY 123 CTGGGCTCCCTCTCTCTGCGCGCGCGCGCGCGCTGCGCGCAAGCAGCAGCTCGTGAC 182  
 DB 70 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAlaValAlaAsp 89  
 QY 183 CTGACCTTCTTCACCCAGAGCGCTGCACTGCTGACCCCTGCTCTGCTGCTCACCC 242  
 DB 90 LeuAepPhePheThrGlnGluProLeuHisLeuValSerProSerPheLeuSerValThr 109  
 QY 243 ATTGACGCCAACCTGCGCGCGCGCGCGCGCTTCCATCTCTCTGCTGCTTCCAAAGCTT 302  
 DB 110 IleAepAlaAenLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProLysLeu 129  
 QY 303 CGTACCTTGGCCAGAGGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362  
 DB 130 ArgThrLeuAlaArgGlyLeuSerProAlaTyrLeuAlaArgPheGlyTyrThrLysThrAsp 149  
 QY 363 TTCTTAATTTTCGATCCCAAGAGGAATCAACTTTGAAGAGAGATTACTGCGCAATCT 422  
 DB 150 PheLeuIlePheAepProLysGlyLeuSerThrPheGlnGluArgSerTyrTrpGlnSer 169  
 QY 423 CAAGTCACACGAGATTTTGCAGAAATATGATTCATCTCTCTGATGCTGAGAGAGATTA 482  
 DB 170 GlnValaAenGlnAaspIleCysLysTyrGlySerIleProProAepValGlnGluLysLeu 189  
 QY 483 CGGTGAATGGCCCTTACAGAGCAATTTGCTACTCCGAGAACATCACAGAAAGTTTC 542  
 DB 190 ArgLeuGlnTrpProTyrGlnGlnGlnLeuLeuAArgGlnHisTyrGlnLysSerPhe 209  
 QY 543 AAGAACAGACCTTACTCAAGAAAGCTCTGTAGAATGCTATACATTTTGCAAATGCTCTCA 602  
 DB 210 LysAenSerThrTyrSerArgSerSerValAepValLeuTyrThrPheAlaAenCysSer 229  
 QY 603 GGACTGCACTGATCTTTGGCTTAATGCGTTATTAAGAACAGCAGATTGAGAGTGAAC 662  
 DB 220 GlyLeuAepLeuIlePheGlyLeuAenAlaLeuLeuAArgThrAlaAepLeuGlnTrpAsn 249  
 QY 663 AGTTCAATGCTCAGTTGCTCTGAGTACTGCTCTTCAAGGGGTATTAACATTTCTTGG 722  
 DB 250 SerSerAenAlaGlnLeuLeuLeuAaspTyrCysSerSerLysGlyTyrAenIleSerTrp 269  
 QY 723 GAAGTGAAGCAATGAACCTTAACAGTTTCTTAAGAGGCTGATATTTTCATCAATGGTTCG 782  
 DB 270 GlnLeuGlyAenGlnProAenSerPheLeuLysLysAlaAepIlePheIleAenGlySer 289  
 QY 783 CAGTTAGAGAAAGATTATTAATTCATTAATGCAATAAATCTTAAGAAATCCACCTTCAAAAT 842  
 DB 290 GlnLeuGlyGlyAaspTyrIleGlnLeuHisLysLeuLeuAArgLysSerThrPheLysAsn 309  
 QY 843 GCAAAATCTATGATGCTGATGTTGGTGCAGCTCGAAGAAAGAGCGCTTAAGTGCAGAG 902  
 DB 310 AlaLysLeuTyrGlyProAepValGlyGlnProAArgLysThrAlaLysSerLeuLys 329  
 QY 903 AGCTTCTGAAGGCTGCTGAGAGAACTGATTAATTCATTAATGCAATCACTACTATTTTG 962  
 DB 330 SerPheLeuLysAlaGlyGlyGluValIleAepSerValIleThrPheHisTyrTrpLeu 349  
 QY 963 AATGACGAGACTGCTTACAGGAGAGATTTTCTTAAACCTGATGTAATTGACATTTTATTT 1022  
 DB 350 AenGlyAArgThrAlaThrArgGluAaspPheLeuAenProAepValLeuAaspIlePheIle 369  
 QY 1023 TCATCTGTCGCAAAAGTTTTCAGGTGTTGAGAGCAGCGCTGCGCAAGAGGCTTGG 1082  
 DB 370 SerSerValGlnLysValPheGlnValAlaGlnSerThrAArgProGlyLysLysValTrp 389  
 QY 1083 TTAGAGAAACAAGCTTGCATATATGAGCGAGAGCGCCCTTGTCTATCCGACATTTTGA 1142

```

Db      390 LeuGIyGIuThrSerSerAlaTyRGIyGIyAlaProIeuLeuSerAapThrPheAla 409
QY      1143 GCTGGCTTTATGTTGGTGAATTAATTTGGGCTGTCAAGCCGAAATGAAATGAGTGTG 1202
Db      410 AlaGlyHehSerTrpLeuSerLeuGIyLeuSerAlaArgMetGIyIleGIyValVal 429
QY      1203 ATGAGGCAAGTATCTTTGGAGCAGAACTACCATTTAGTGAATGAATTCGATCTCT 1262
Db      430 MetArgGIuValPhePheGIyAlaGIyAsnTyRHisLeuValAArgGIuAsnPheAapPro 449
QY      1263 TTACCGTATATTTGGCATCTCTCTTTCAAGAAATTTGGGGCCACCAAGCTTTATG 1332
Db      450 LeuProAapTyRTrpLeuSerLeuLeuPheLeuValGIyTrpValLeuMet 469
QY      1323 GCAAGCTGCAAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382
Db      470 AlaSerValGIuGIySerTyRArgRGIyLeuArgValTyRLeuHisCysThrAsnThr 489
QY      1383 GACAAATCCAAAGTATTAAGAGAGAGATTTAACTGTATGCCATTAACCTCCATACGTC 1442
Db      490 AspAsnProArgTyRGIySGIuGIyAspLeuThrLeuTyRAlaIleAsnLeuHisAsnVal 509
QY      1443 ACCAAGACTTGGGGTTACCTATCCCTTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 1502
Db      510 ThrIysTyRLeuArgLeuProTyRProPheSerAsnIySGIuValAapIySGIyLeuLeu 529
QY      1503 AGACCTTTGGAGACCTCATGATTAATCTTTCCAAATCTGTCCAACTCAATGCTCAACTCTA 1562
Db      530 ArgProLeuGIyProHisGIyLeuLeuSerIySerValGIuLeuAsnGIyLeuThrLeu 549
QY      1563 AAGATGTTGATGATCAAACTTCCCACTTTAATGAAAAAATCTTCCGGCCAGGAAGT 1622
Db      550 LysMetValAapArgGIuTrpLeuProLeuMetGIuIySGIyProLeuArgProGIySer 569
QY      1623 TCACCTGGGCTTGGCAGCTTCTCATATAGTTTTTTTGGATTAAGAAATGCCAAAGTGGCT 1682
Db      570 SerLeuGIyLeuProAlaPheSerTyRSerPhePheValIleArgAsnAlaIyValAla 589
QY      1683 GCTTGCATC 1691
Db      590 AlaCysIle 592

RESULT 3
AAI30124
ID      AAY30124 standard; protein; 588 AA.
XX
XX      AAY30124;
XX
XX      20-MAR-2003 (revised)
XX      14-OCT-1999 (first entry)
XX
XX      A human protein with heparanase activity.
XX
XX      Human; heparanase; heparan sulfatase; trauma; autoimmune disease;
XX      skin disease; cardiovascular disease; nervous system disease;
XX      Alzheimer's disease; cancer; cancer metastasis; angiogenesis;
XX      inflammation; arthritis.
XX
XX      Homo sapiens.
XX
XX      MO9940207-A1.
XX
XX      12-AUG-1999.
XX
XX      05-FEB-1999; 99MO-EP000777.
XX
XX      09-FEB-1998; 98GB-000002725.
XX
XX      (NOVS ) NOVARTIS AG.
XX      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX      Nakajima M, Toyoshima M;

```

```

DR      WPI; 1999-494300/41.
DR      N-PSDB; AAX86671.
XX
XX      New heparanase polypeptide useful for treating autoimmune diseases, skin
PT      diseases, cardiovascular diseases and nervous system diseases including
PT      Alzheimer's disease.
PT      Claim 3; Page 29-31; 40pp; English.
XX
XX      The present sequence represents a polypeptide with human heparanase
CC      biological activity. Antagonists and inhibitors of the protein prevent it
CC      from degrading the extracellular matrix and releasing heparan sulfate
CC      from the extracellular matrix surface. The heparanase protein or the anti
CC      -heparanase antibody are used in pharmaceutical compositions for treating
CC      warm blooded animals suffering from a disease resulting from shortage or
CC      lack of the heparanase protein, or from excessive activity or over-
CC      expression of the heparanase protein, respectively. The heparanase
CC      protein is used in treating diseases such as trauma, autoimmune disease,
CC      skin diseases, cardiovascular diseases and nervous system diseases
CC      including Alzheimer's disease resulting from shortage or lack of
CC      polypeptide. The anti-heparanase antibody is used in treating the
CC      diseases like cancer, cancer metastasis, angiogenesis and inflammation
CC      including arthritis resulting from excessive activity or over expression
CC      of heparanase protein. The anti-heparanase antibody can be used to detect
CC      the presence or absence of polypeptide and its concentration. (Updated on
CC      20-MAR-2003 to correct PA field.)
XX
XX      Sequence 588 AA;
SQ
XX
XX      Alignment Scores:
XX      Pred. No: 8,756-279 Length: 588
XX      Score: 2904.00 Matches: 553
XX      Percent Similarity: 100.00% Conservative: 1
XX      Best Local Similarity: 99.82% Mismatches: 0
XX      Query Match: 93.11% Indels: 0
XX      DB: 2 Gaps: 0

US-10-676-079-3 (1-1721) x AAY30124 (1-588)
QY      30 AGCTGGGGGGGAGGAGCAGGAGTGAAGCCCAAGTCTGTGGCTGAAAGCTGGGCGT 89
Db      35 SerTrpArgGIyGIuGIuTrpGIyGIuProIyMetLeuLeuArgSerIyProAlaLeu 54
QY      90 CCGCCGCGCTGATGCTGCTCTCTGGGGCGCTGGGCTCCCTCTCCCTGGGCGCTTG 149
Db      55 ProProProLeuMetLeuLeuLeuLeuGIyProLeuGIyProLeuSerProGIyAlaLeu 74
QY      150 CCGGACCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 209
Db      75 ProArgProAlaGIuIleAsnIleAsnValIleAspLeuAspPhePheThrGIuIleProLeu 94
QY      210 CACCTGGTGAAGCCCTCGTTCCTGCTCCGTCACCATTTGAGCCCAAGCTGGCCAGAGCCG 269
Db      95 HisLeuValSerProSerPheLeuSerValThrIleAspIleAsnLeuAlaThrAapPro 114
QY      270 CGGTTCTCATCTCTCTGGGTTCTTCAAAAGCTTGTAACCTTGGCCAGAGGCTTGTCTCT 329
Db      115 ArgPheLeuIleLeuLeuGIySerProIyLeuArgThrLeuAlaArgGIyLeuSerPro 134
QY      330 GCGTACCTGAGGTTTGGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 389
Db      135 AlaTyRLeuArgPheGIyGIyTrpIyAspPheLeuIlePheAspProIySGIu 154
QY      390 TCAACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 449
Db      155 SerThrPheGIuIleuArgSerTyRTrpGIuSerGIuValAsnIleAsnIleCysIyTyR 174
QY      450 GATTCATCTCTCTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 509
Db      175 GlySerIleProProAspValGIuGIuIyLeuArgLeuGIuTrpProTyRGIuGIuGIu 194
QY      510 TTGCTACTCCGAGAACTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 569

```



```

Db      195  LeuLeuLeuArgGluHleTyrGlnLysLysPheLysAasnSerThrTyrSerArgSerSer 214
Qy      570  GAGATGAGCTATTAACCTTTGGCAACGCTGAGGACCTGATCTTGGCCCTTAAT 629
Db      215  ValAspValLeuTyrThrPheAlaAsnCySerGlyLeuAspLeuIlePheIleLysAsn 234
Qy      630  GCGTATTAAAGAACAGATTGTCAGTGAACAGTTCTTAAGTCTCAGTTGCTCTGGAC 689
Db      235  AlaLeuLeuArgThrAlaAspLeuGlnTrpAsnSerSerAsnAlaGlnLeuLeuLeuAsp 254
Qy      690  TACTGCTCTTCCAGGGGTATTAACATTTCTGGGAACCTAGCAATGAACCTACAGTTTC 749
Db      255  TyrCySerSerSerLysGlyTyrAsnIleSerTrpGlnLeuGlyAsnGlnProAsnSerPhe 274
Qy      750  CTTAAGAAGCGCTGATATTTCATCAATGGTGGCGATTAGAGGAAGATTAATTCGAATTG 809
Db      275  LeuLysLysAlaAspIlePheIleAsnGlySerGlnLeuGlyGlnAspPheIleGlnLeu 294
Qy      810  CATAAACCTTCAAGAAAGTCACCTTCAAAAATGCAAAACTTATAGTCTGATGTTGGT 869
Db      295  HisLysLeuLeuArgLysSerThrPheLysAsnAlaLysLeuTyrGlyProAspValGly 314
Qy      870  CAGCTCGAAGAAGACGGCTAAGATGCTGAAGAGCTTCGAAGGCTGTCGAGAAGTG 929
Db      315  GlnProArgArgLysThrAlaLysMetLeuLysSerPheLeuLysAlaGlyGlnVal 334
Qy      930  ATTGATTGATTAACATGGATCACTATTTGATGACGGAACGCTACCAAGGAAGAT 989
Db      335  IleAspSerValThrTrpHisIleTyrLysLeuAsnGlyArgThrAlaThrArgLysAsp 354
Qy      990  TTTCTAAACCTGATGATTTGACATTTTATTTTCAATCTGTCGCAAAAAGTTTCCAGGTG 1049
Db      355  PheLeuAsnProAspValLeuAspIlePheIleSerSerValGlnLysValPheGlnVal 374
Qy      1050  GTTGAAGACACCAAGGCTTGGCAAGAGGTCTGGTTAGCAAAACAAGCTTCATATGGA 1109
Db      375  ValGlnSerThrArgProGlyLysLysValTrpLeuGlyGlnThrSerSerAlaTyrGly 394
Qy      1110  GGGGAGAGCGGCTTGATCCGACACCTTTGACAGGCGCTTATGCGGTGATTAATG 1169
Db      395  GlyGlyAlaProLeuLeuSerAspThrPheAlaAlaGlyPheMetTrpLeuAspLysLeu 414
Qy      1170  GGCCTGTCAAGCCGAATGGAATAGAAAGTGATGAGGCAAGTATCTTGGACAGGA 1229
Db      415  GlyLeuSerAlaArgMetCylIleGlnValValMetArgGlnValPhePheGlyAlaGly 434
Qy      1230  AACTACATTTAGTGATGAAAACTTGATCTTAACTGATTAATGCTATCTCTTG 1289
Db      435  AsnTyrHisLeuValAspGlnAsnPheAspProLeuProAspTyrTrpLeuSerLeuLeu 454
Qy      1290  TTCAGAAATTTGGTGGGACCAAGGTGTTAATGGCAAGGTCGCAAGGTTCAAGAGAG 1349
Db      455  PheLysLysLeuValGlyThrLysValLeuMetAlaSerValGlnGlySerLysArgAsp 474
Qy      1350  AAGCTTCGAGTATTAACCTTCATTCGACAAACACTGCAATCCAAAGGTATTAAGAGAGAT 1409
Db      475  LysLeuAsnGlyValTyrLeuHisCysThrAsnThrAspAsnProArgTyrLysGlnGlyAsp 494
Qy      1410  TTAACCTGTATGCCATAAACCCTCAATACGTCACCAAGTACTTGGGTTACCTTACCT 1469
Db      495  LeuThrLeuTyrAlaIleAsnLeuHisAsnValThrLysTyrLeuArgLeuProTyrPro 514
Qy      1470  TTTTCTTAAGACAGATGATTAATACCTTTAAAGCCTTTGGACCTCAATGATTACTT 1529
Db      515  PheSerAsnLysGlnValAspLysTyrLeuLeuAspProLeuGlyProHisGlyLeuLeu 534
Qy      1530  TCCAAATCTGTCAACATCAATGCTTAACTTAAAGATGGATGATGAAAACCTTGCCA 1589
Db      535  SerLysSerValGlnLeuAsnGlyLeuThrLeuLysMetValAspAspGlnThrLeuPro 554
Qy      1590  CCTTAAATGAAAAACCTCTCGGCGCAAGAGTTCAGTGGGCTTGCGACCTTCTCATAT 1649
Db      555  ProLeuMetGlnLysProLeuArgProGlySerSerLeuGlyLeuProAlaPheSerTyr 574

```

```

Qy      1650  AGTTTGTGATTAAGAAATGCCAAGTTCCTGCTGCATC 1691
Db      575  SerPhePheValIleArgAsnAlaLysValAlaAlaCysIle 588

```

## RESULT 4

AAV02345

ID AAV02345 standard; protein; 543 AA.

AC AAV02345;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

US-10-676-079-3 (1-1721) x AAV02345 (1-543)

```

Sequence 543 AA;

Alignment Scores:

Pred. No.: 1,2e-272

Score: 2842.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 91.12%

DB: 2

Gaps: 0

Length: 543

Matches: 543

Conservative: 0

Mismatch: 0

Indels: 0

Gaps: 0

New human polynucleotide useful for treating angiogenesis, restenosis, and inflammation.

Claim 6; Fig 1; 63pp; English.

The specification describes a polypeptide having heparanase (hp) activity. The recombinant protein is used as a modulator of heparin-binding growth factors, cellular responses to heparin-binding growth factors and cytokines, cell interaction with plasma lipoproteins, cellular susceptibility to viral, protozoal and bacterial infections or conditions such as wound healing, angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases, and viral infections. Mammalian heparanase can be used to neutralize plasma heparin, and anti-heparanase antibodies may be applied for immunodetection and diagnosis of micrometastases, autoimmune lesions, and renal failure in biopsy specimens, plasma samples, and body fluids. The present sequence represents human heparanase

QY	63	ATGCTGCTGCGCTCGAAGCGCTGCGCGCGCGCTGAATGCTGCTGCTGGGGCGCG	122
Db	1	MetLeuLeuAaGSerLysProAlaLeuProProProLeuMetLeuLeuLeuLeuGlyPro	20
QY	123	CTGGGTCCTCTCTCCCTGGCGCCCTGGCCGACCTGGCGAAGCAGCAGGACGCTGGAC	182
Db	21	LeuGlyProLeuSerProGlyAlaLeuProAlaGlnAlaGlnAlaValAlaLeu	40
QY	183	CTGGAACTTTTCAACCGAGAGCCGCTGCACCTGGTGAGGCCCTCGTTCTGTGCTGAC	242
Db	41	LeuAaPhePheThrGlnIuProLeuH1sLeuValSerProSerPheLeuSerValThr	60
QY	243	ATTGACGCCAAGCTGGCGCACGGGACCGCGGCTTCATCTCTCGGGGTTCTCGAAGCTT	302
Db	61	IleAaPheAlaMetLeuAlaThrAaProAaGlnPheLeuI1LeuLeuGlySerProLysLeu	80
QY	303	CGTACCTTGGCCAGAGGCTTGTCTCTGGCTACCTGAGGTTTGGTGACACAGACAGAC	362
Db	81	ArgThrLeuAlaArgGlyLeuSerProAlaTyRLeuAaGlnPheGlyGlyThrLysThrAaP	100
QY	363	TTCTTAATTTTGGATGCCAAGAGGATACACCTTTGAAGAGAAAGTTATCTGCAATCT	422
Db	101	PheLeuI1ePheAaPProLysArgLysGlnSerThrPheGlnGlnAaGSerTyTrpGlnSer	120
QY	423	CAAGTCACACCGGATATTGGCAAAATATGATGATCCATCCCTCGAGTGAGAGAGAAAGTTA	482
Db	121	GlnValAaenGlnAaP1eCysLysTyRGlSerI1eProProAaPValGlnIuLysLeu	140
QY	483	CGGTATGGAATGGCCCTTACCGAGAGCAATTGCTACTCCGAGAACATCAACAGAAAAAGTTC	542
Db	141	ArgLeuGlnTrpProTyRGlndIuGlnLeuLeuLeuAaArgGlnH1sTyRGlndLysSerPhe	160
QY	543	AAGAACAGACCTTACTCAAGAAAGCTCTGTAGATGGCTATACACTTTGGCAATGCTCA	602
Db	161	LysAaenSerThrTySerArgSerSerValAaPValLeuTyTrhPheAlaAaenCysSer	180
QY	603	GGACCTGGAAGCTTGGAGGCTTAATGGCTAATTAAGAACAAGCATTTGGACGTGAGAC	662
Db	181	GlyLeuAaPheLeuI1ePheGlyLeuAaenAlaLeuLeuAaGlyThrAlaAaPLeuGlnTrpAaen	200
QY	663	AGTTTAATGCTCAAGTGTCTCTGGACTACTCTCTTCCAAAGGGATATTAACATTTCTTGG	722
Db	201	SerSerAaenAlaGlnLeuLeuLeuAaenPyrCysSerSerLysGlyTyRAsnI1eSerTrp	220
QY	723	GAATGAGCAATGAACCTTACAGTTTCTTTAAGAAAGCTGATATTTTCATCAATGGGTCG	782
Db	221	GlnLeuGlyAaenGlnProAaenSerPheLeuLysValAaP1ePheI1eAaenGlySer	240
QY	783	CAGTTAGAGAGAATTAATTAATTCATATGGCATAACTTTGAAGAAATCCACCTTCAAAAT	842
Db	241	GlnLeuGlyGlnAaPTrpL1eGlnLeuH1sLysLeuLeuAaGlyLysSerThrPheLysAaen	260
QY	843	GCAAAACTCTATGCTGCTGATTTGGTGTGAGCCTCGAAGAAAGACGGCTGAAGTGGTAGAG	902
Db	261	AlaLysLeuTyRGlYProAaPValGlyGlnProAaGlnPValstH1AlaLysMetLeuLys	280
QY	903	AGCTTCTGAAGGCTGGTGAGAAAGATGATGATCAGTTACATGGCATCACTACTATTTGG	962
Db	281	SerPheLeuLysAlaGlyGlyGlyuValI1leAaPSerValThrTrpH1sH1sTyRTrLeu	300
QY	963	AATGACCGGATGCTTACCGAGGGAAGATTTTCTAAACCTGATGlnATTGGACATTTTATTT	1022
Db	301	AaenGlyAaGlyThrAlaThrArgGlnAaPheLeuAaenProAaPValLeuAaP1ePheI1e	320
QY	1023	TCATCTGTCGCAAAAAGTTTCCAGGTGGTGAGACACAGGCGTGGGGAAGAAAGCTTGG	1082
Db	321	SerSerValGlnLysValPheGlnAlaValGlnSerThrArgProGlyLysLysValAlaTrp	340
QY	1083	TTAGAGAGAAACAAGCTCTGCATATATGAGAGCGAGCGCCTTGTCTATCCGACACCTTTGCA	1142
Db	341	LeuGlyGlnTrpSerSerAlaTyRGlYglYglAlaProLeuLeuSerAaPThrPheAla	360
QY	1143	GCTGGCTTATATGTGCTGGATTAATTTGGGCGCTGTGACGCCGAGTGGAAATGAAGTGGT	1202

Db	361	AlaGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaAspGlyIleGlnValVal	380
QY	1203	ATGAGGCAAGTATCTTTGGACAGAAACTACCATTTAGTGATGAAGAACTTCGATCT	1262
Db	381	MetArgGlnValPhePheGlyAlaGlyAsnTyrHisLeuValAspGlnAsnPheAspPro	400
QY	1283	TTACTGATTTATTTGGCTATCTCTTGTGTTCAAGAAATTTGGTGGGCAACAGTGTTAATG	1322
Db	401	LeuProAspTyrTrpLeuSerLeuIleuPheLysLeuValGlyTyrHisValLeuMet	420
QY	1323	GCAAGCGTGCAGAGCTTCAAGAGAGAAAGCTTCGAGTATACCTTCATTTGCACAAACACT	1382
Db	421	AlaSerValGlnGlySerLysValArgValLysLeuArgValTyrLeuHisCysThrAsnThr	440
QY	1383	GACAAATCCAAAGTATTAAGAGAGATTTAACTGTATGTCATTAACCTTCATTAACGTC	1442
Db	441	AspAsnProAspGlyTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLeuHisAsnVal	460
QY	1443	ACCAAGTATCTTGGCGGTTACCTTATCTTTTCTTAACAAGCAAGTGATTAATACCTTCTA	1502
Db	461	ThrTyrTyrLeuAspArgLeuProTyrProPheSerAsnLysGlnValAspLysTyrLeuLeu	480
QY	1503	AGACCTTTGGGAGCCTCATGATGATTACTTTCCAAATCTGGCCCAACAGTGTCTTAACCTTA	1562
Db	481	ArgProLeuGlyProHisLysLysLeuSerLysSerValGlnLeuAspGlyLeuThrLeu	500
QY	1563	AAGATGTGTGATGATCAAAACCTTGCACACTTTAATGCAAAAACCTTCTCGGCGCAGGAAGT	1622
Db	501	LysMetValAspAspArgInThrLeuProIleuMetGlnLysProLeuAspProGlySer	520
QY	1623	TCACCTGGGCTTGGCCAGCTTCTCATATATGTTTTTTGTGATTAAGAAATGCCAAAGTTGCT	1682
Db	521	SerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLysValAla	540
QY	1683	GCTTGCAATC 1691	
Db	541	AlaCysIle 543	
RESULT 5			
ID	AAV57590	AAV57590 standard; protein; 543 AA.	
XX	AAV57590;		
DT	02-MAR-2000 (first entry)		
XX	DE	Human heparanase.	
XX	Human, heparanase; hpa; genetic modification; expression; anticancer;		
KW	angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumor;		
KW	anti-atherosclerotic; anti-inflammatory; antineurodegeneration;		
KW	heparan sulphate; heparin-binding growth factor; tumor angiogenesis;		
KW	metastasis; wound healing; restenosis; atherosclerosis; inflammation;		
KW	neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;		
KW	micrometastasis; autoimmune lesion; kidney failure.		
XX	Homo sapiens.		
XX	OS		
XX	PN	W09957244-A1.	
XX	PD	11-NOV-1999.	
XX	PF	29-APR-1999; 99WO-US009256.	
XX	PR	01-MAY-1998; 98US-00071618.	
XX	PR	02-MAR-1999; 99US-00260038.	
PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.		
XX	(PRIE/) FRIEDMAN M M.		
XX	Ben-Attzi H, Ayal-Herehkovitz M, Yacoby-Zeevi O, Becker I;		
XX	Peleg Y, Shlom Y,		

XX MPI: 2000-062144/05.  
DR N-PSDB; AA239195.

PT Engineered cells that express recombinant heparanase, useful  
PT therapeutically, e.g. for treating angiogenesis and to screen for  
XX specific inhibitors, potential anticancer agents.

PS Claim 3; Page 107-109; 118pp; English.

XX The present invention describes genetically modified cells (A) containing  
CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,  
CC and express recombinant heparanase (II). Heparanase cleaves heparan  
CC sulphate (HS) at specific intrachain sites, resulting in release of  
CC heparin-binding growth factors, enzymes and proteins that are sequestered  
CC by HS in basement membranes, extracellular matrix or cell surfaces. It  
CC may also be implicated in tumour angiogenesis and metastases. (II) is  
CC potentially useful in wound healing and for treating angiogenesis,  
CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral  
CC infection and cystic fibrosis. It can also be used to neutralise heparin  
CC (an alternative to protamine) and to screen for specific inhibitors  
CC (potentially useful for treating cancer and metastases). Antibodies  
CC raised against (II) are used for immunodetection and diagnosis of  
CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)  
CC in large quantities, in a form that is homogeneously processed and  
CC activated/neutralised by a dedicated protease. The present sequence  
CC represents human heparanase

XX Sequence 543 AA;

Alignment Scores:

Pred. No.:	1-2e-272	Length:	543
Score:	2862.00	Matches:	543
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	91.12%	Indels:	0
DB:	3	Gaps:	0

US-10-676-079-3 (1-1721) x AAY57590 (1-543)

QY 63 ATGCTGCTGCGCTGAGAGCGCTGCGCCGCGCTGATGCTGCTGCTGCGGCGG 122  
DB 1 MetLeuLeuArgSerIysProAlaLeuProProProLeuMetLeuLeuLeuGlyPro 20  
QY 123 CTGGGCTCCCTCTCCCTGCGCCGCTGCGCCGCTGCGGAGCAAGCAAGCAAGCA 182  
DB 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaInspValAlaSer 40  
QY 183 CTGGACTTCTTCAACGAGAGCGCTGCACTGCTGAGGCGCTGCTGCTGCTGCTGCTG 242  
DB 41 LeuAspPheMetThrGlnGluProLeuHisLeuValSerProSerPheLeuSerValThr 60  
QY 243 ATTGACGCAACCTGCGCAGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302  
DB 61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProLysLeu 80  
QY 303 CGTACCTTGGCCAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362  
DB 81 ArgThrLeuAlaArgGlyLeuSerProAlaTyLeuAlaArgPheGlyGlyThrIlyStrAsp 100  
QY 363 TTCCTAATTGATCCCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 422  
DB 101 PheLeuIlePheAspProLysLysGlySerThrPheGlnGlnArgSerIyTrpGlnSer 120  
QY 423 CAAGTCAACAGAGATATTTGCAATATGATTCATCTCTGATGCTGAGAGCAAGTGA 482  
DB 121 GlnValAsnGlnInspIleCyIySlyrGlySerIleProProAspValGlnGlnLysLeu 140  
QY 483 CGGTGGATGCGCCCTTCAACAGAGCAAGTCTTCTGAGAGCAAGCAAGCAAGCAAG 542  
DB 141 ArgLeuGlnIlyrProIyGlnGlnGlnLeuLeuArgGlyIlyrGlnLysLysPhe 160  
QY 543 AAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 602

DB 161 LysAsnSerThrTySerArgSerSerValAspValLeuTyThrPheAlaAsnCysSer 180  
QY 603 GACTGCACTTGATCTTTGGCTTAAGCCCTTATTAAGACAGAGATTTGACATGGAGAC 662  
DB 181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn 200  
QY 663 AGTTCTAATGCTCAGTTGCTCCTGAGCACTGCTCTTCCAGAGGATTAACATTTCTTGG 722  
DB 201 SerSerAsnAlaGlnLeuLeuLeuAspTyTySerSerSerIySgIyTyAsnIleSerTrp 220  
QY 723 GAATGAGCAATGAACCTTAACAGTTTCTTAAGAGGCTGATATTTTCATCATGAGGCTG 782  
DB 221 GlnLeuGlyAsnGlnProAsnSerPheLeuLysValAspIlePheIleAsnGlySer 240  
QY 783 CAGTTAGAGAGAGATTAATTCATTTGATTAACCTTCAAGAAAGTCCACTTCAAAAT 842  
DB 241 GlnLeuGlyGlnAspTyTyrIleGlnLeuHisLysLeuLeuArgLysSerThrPheLysAsn 260  
QY 843 GCAAAACTCTATGCTGCTGATGTTGCTGAGCCTGGAAGAAAGAGGCTTAAGATGCTGAG 902  
DB 261 AlaLysLeuTyTyrGlyProAspValGlyGlnProArgGlyStrAlaLysMetLeuLys 280  
QY 903 AGCTTCTGAAGGCTGAGTGAAGAACTGATGATTCATGATGATGATGATGATGATGATG 962  
DB 281 SerPheLeuLysAlaGlyGlyGlyValIleAspSerValThrTrpHisIleTyTyTrpLeu 300  
QY 963 AATGACGCACTGCTGCTCAACAGAGAAATTTTCTAAACCTGATGATGATGATGATGAT 1022  
DB 301 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320  
QY 1023 TCATCTGTCAAAAGTTTTCAGAGTGTGAGAGCAGAGCCTGCAAGAGAGGCTGAG 1082  
DB 321 SerSerValGlnLysValPheGlnValAlaGlnSerThrArgProGlyLysLysValTrp 340  
QY 1083 TTAGAGAAACAAGCTCTGCAATATGAGAGCGAGGCGCTTGTCTATCCGACACCTTTGCA 1142  
DB 341 LeuGlyGlnThrSerSerAlaTyTyGlyGlyAlaProLeuLeuSerAspThrPheAla 360  
QY 1143 GCTGCTTATATGCTGCTGATTAATTTGGGCTGTGACGCCGCAATGGAATGAAAGTGTG 1202  
DB 361 AlaGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaAspGlyIleGlnValAla 380  
QY 1203 ATGAGCAAGATTTCTTTGAGCAGAGAAACTACATTAATGAGATGAAGAACTTCAGACT 1262  
DB 381 MetArgGlnValPhePheGlyAlaGlyAsnTyThrIleValAspGlnAspPheAspPro 400  
QY 1263 TTACTGATTAATTTGCTATCTCTTCTGTTCAAGAAATTTGTTGGGACCAAGGTTTAA 1322  
DB 401 LeuProAspTyTrpLeuSerIleuLeuPheLysLysValGlyThrIlyValLeuMet 420  
QY 1323 GCAAGCGTCAAGGTTCAAGAGAGAAAGAACTTCAAGATATCTTCAATTTGCAAAACCT 1382  
DB 421 AlaSerValGlnGlySerLysArgIlyLysLeuArgValTyTyrLeuHisCysThrAsnThr 440  
QY 1383 GACATTCAGAGTAAAGAGAGAGATTAACCTGATGACATTAACCTGATTAACCTGAT 1442  
DB 441 AspAsnProArgTyTyrLeuGlnGlyAspLeuThrLeuTyAlaIleAsnLeuHisAsnVal 460  
QY 1443 ACCAAGTCTTGGCTTACCTTATCTTCTTCTTCAAGAGCAAGTGAATTAATCTTCTTA 1502  
DB 461 ThrIlyTyTyrLeuArgLeuProTyTrpPheSerAsnLysGlnValAspLysTyTyrLeu 480  
QY 1503 AGACTTTGGAGCTCAGATGATTAATCTTCAATCTTGTCAACTCAATGCTTAATCTCTA 1562  
DB 481 ArgProLeuGlyProIleGlyLeuLeuSerIySerValGlnLeuAsnGlyLeuThrIleu 500  
QY 1563 AAGATGCTGAGATGAACAACTTGGCACTTAAAGGAAAGAACTTCGAGCAGAGAGT 1622  
DB 501 LysMetValAspAspGlnThrIleuProProLeuMetGlnLysPheLeuArgProLysSer 520  
QY 1623 TCATGCGCTGCGCAGCTTCTCATATAGTTTCTTGTGATGAAGAAATGCCAAGTGTCT 1682



QY 1143 GCTGGCTTATGTCGTCGATAAATGCGCTGACCCGAGTGGATAGAGTGTG 1202  
 DB 361 AAGAGlyPheMetITripleuAspIlyLeuSerAlaIrgMeGlyIleIValVal 380  
 QY 1203 ATGAGGCAAGATATCTTTGGACGACGAAATACCATTTAGTGGATGAAAATTGCATCT 1262  
 DB 381 MetArgGlnValIlePheMetIlyAlaGlyAenITyHileValIAspGluAenPheAspPro 400  
 QY 1263 TTACCTGATTTATGGTATCTCTGTCGTCGAAATGGTGGGACCAAGGTGTTAAG 1322  
 DB 401 LeuProAspTyrITriPheuSerLeuPheIlyLeuValGlyThrllySValIleuMet 420  
 QY 1323 GCAAGCGTCAGAGTTCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382  
 DB 421 AlasSerValGlnGlySerIlyAspArgIlyIleuArgValIlyIleuHicSerThrAnthr 440  
 QY 1383 GACATTCAGAGTATTAAG 1442  
 DB 441 AspAspProArgTyrIlySerIlyGluGlyAspLeuThrLeuTyrAlaIleAenIleuHicAsnVal 460  
 QY 1443 ACCAAGTATTTGCGCTTACCTATCTTTTTCGAAACAGAGAGAGAGAGAGAGAGAG 1502  
 DB 461 ThrIlyTyrLeuArgLeuProTyrProPheSerAsnIlyGlnValAspIlyTyrIleuLeu 480  
 QY 1503 AGACCTTTGGAGACCTGATGATTAATCTTCCAAATCTGTCGAACTCAATGATCTTA 1562  
 DB 481 ArgProIleuGlyProHicIlyLeuLeuSerIlySerValGlnIleuAsnIlyLeuThrIleu 500  
 QY 1563 AAGATGTCGATGATCAAACTTTGCCACCTTTAATGAGAAAACCTTCCGGCCAGAGAG 1622  
 DB 501 LysMetValAspAspGlnThrIleuProIleuMetGluIlyProIleuArgProGlySer 520  
 QY 1623 TCACCTGGCTTGGACACTTTCATATAGTTTTCGATATAGAAATCCCAAGTGTCT 1682  
 DB 521 SerLeuGlyLeuProIleuPheSerTyrSerPheValIleArgAsnAlaIlyValAla 540  
 QY 1683 GCTTGCATC 1691  
 DB 541 AlaCysIle 543  
 RESULT 7  
 AAY52990  
 ID AAY52990 standard; protein; 543 AA.  
 XX  
 AC AAY52990;  
 XX  
 DT 21-FEB-2000 (first entry)  
 DE Human heparanase protein sequence.  
 XX  
 KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;  
 KW anti-diabetic; immunomodulatory; anti-inflammatory; nephrotropic;  
 KW metacarcinoma; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;  
 KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;  
 KW inflammation; haemorrhagic nephritis; nephrotic syndrome;  
 KW autoimmune disease; anticancer; kidney disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9957153-A1.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 29-APR-1999; 99MO-US009255.  
 XX  
 PR 01-MAY-1998; 98US-00071739.  
 XX  
 PA (INST-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 PI Pecker I, Vlodevsky I, Friedman Y, Perets T;  
 XX

DR WPI: 2000-052944/04.  
 DR N-PSDB; AAZ33290.  
 XX  
 PT Heparanase-specific molecular probes useful for diagnosis and treatment,  
 PT e.g. of tumors, and for targeted drug delivery.  
 XX  
 XX Example; Page 81-82; 90pp; English.  
 XX  
 CC The present invention describes heparanase-specific molecular probes,  
 CC useful for methods of detecting heparanase in situ. The probes and anti-  
 CC heparanase antibodies are used to detect or quantify the expression of  
 CC heparanase, for diagnosis and monitoring of diseases (especially  
 CC metastases), for treatment of heparanase-associated diseases (e.g.  
 CC tumors, adenocarcinoma, squamous cell carcinoma, teratocarcinoma,  
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its  
 CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,  
 CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,  
 CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,  
 CC sepsis and inflammatory or autoimmune disease), for targeted drug  
 CC delivery (e.g. of anticancer agents) and as research reagents. The  
 CC present sequence represents human heparanase, which is used in the  
 CC exemplification of the present invention  
 CC  
 SQ Sequence 543 AA;  
 XX  
 Alignment Scores:  
 Pred. No.: 1,26-272 Length: 543  
 Score: 2842.00 Matches: 543  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 91.12% Indels: 0  
 DB: 3 Gaps: 0  
 US-10-676-079-3 (1-1721) x AAY52990 (1-543)  
 QY 63 ATGCTGTCGCTGTCGACCTGCGCTGCGCGCGCTGATGCTGCTCTCTGGGCGCG 122  
 DB 1 MetLeuLeuArgSerIlyProAlaIleuProProIleuMetLeuLeuIleuIlyPro 20  
 QY 123 CTGGTCCCTCTTCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182  
 DB 21 LeuGlyProIleuSerProGlyAlaLeuProArgProAlaGlnAlaAspValAlaAsp 40  
 QY 183 CTGACCTCTTCAACCCAGAGCGCGCTGACCTGAGAGCCCTCGTCTCTGCTCCATCAC 242  
 DB 41 LeuAspPhePheThrGlnIlyProIleuHicIleuValSerProSerPheIleuSerValThr 60  
 QY 243 ATTGACGCCAACCTGGCCACGAGACCGCGGTTCTCATCTCTGAGGTTTTCGAAAGCTT 302  
 DB 61 IleAspAlaAsnLeuAlaThrAspProArgPheIleuIleuLeuGlySerProIlyIleu 80  
 QY 303 CGTACCTTGGCCAGAGGCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362  
 DB 81 ArgThrLeuAlaArgGlyLeuSerProAlaTyrLeuArgPheGlyIlyThrIlyThrAsp 100  
 QY 363 TTCTCAATTTTGGATCCCAAGAGAGATCACTTGAAGAGAGAGAGAGAGAGAGAGAG 422  
 DB 101 PheLeuIlePheAspProIlySerIlySerIlyThrPheGluGluArgSerTyrITriPheIleu 120  
 QY 423 CAAGTCAACAGAGATATTTGCAATATGATTCATCTCTGATGTGGAGAGAGAGAGAG 482  
 DB 121 GlnValAsnGlnAspIleCysIlySerIlyGlySerIleProProAspValGluIlyIleu 140  
 QY 483 CGGTGGAATGGCCCTTACAGAGAGCAATGCTACTCCAGAGAGAGAGAGAGAGAGAG 542  
 DB 141 ArgLeuGluITriProTyrGlnIlyGlnIleuLeuAlaArgGluHicIlyIlyIlySAsp 160  
 QY 543 AAGACAGACCTACTCAAGAGAGCTCTGAGAGTGTATACATCTTTGCAAGAGAGAGAG 602  
 DB 161 LysAsnSerThrIlySerIlySerIlySerIlyValIleValIleuTyrThrPheAlaAsnIlySer 180  
 QY 603 GGACTGACTTATCTTTGGCTTAAATGCTTATTAAGACAGAGAGAGAGAGAGAGAGAG 662

```

Db      181 G|YLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuIleGThrAlaAspLeuGIntTPAsn 200
QY      663 AGTTCATAGCTCAGTTCCTGCTGAGTACTGCTCTTCCAAAGGGATTAACATTTCTTGG 722
Db      201 SerSerSerAsnAlaGlnLeuLeuLeuLeuAspTyrCysSerSerTyrSglYrAsnIleSerTyr 220
QY      723 GAACATGAGCAATGAACTTAACAGTTTCCTTAAGAGGCTGATATTTTCATCATAGGCTCG 782
Db      221 GlnLeuGlyAsnGlnProAsnSerPheLeuSylAspIlePheIleAsnGlySer 240
QY      783 CAGTTAGAGAGAGATTTATTTATTCATTTGATTAACCTTCTAAGAAAGTCCACCTTCAAAAT 842
Db      241 GlnLeuGlyGlnAspTyrIleGlnLeuHisLeuLeuLeuGlySerThrPheLeuAsn 260
QY      843 GCAAACTATAGTCTGATGTTGTGTGACGCTCGAAGAAAGAGGCTTAAGATCTGAAG 902
Db      261 AlaLeuLeuTyrGlyProAspValGlyGlnProArgTyrSglYrAlaLeuMetLeuLys 280
QY      903 AGCTTCCTGAAGGCTGTGAGAAAGTGAATTGATTCATTCATGAGCATCACTACTATTGG 962
Db      281 SerPheLeuSylAspIleGlyGlnValIleAspSerValThrTPHisIleTyrTyrLeu 300
QY      963 AATGAGAGGAGCTGTACAGGAGAGATTTTCTAAACCTGATGATTAATGACATTTTATT 1022
Db      301 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320
QY      1023 TCATCTGTGCAAAAAGTTTTCAGAGTGTGAGAGCAACAGGCTGCGCAAGAGTCTGG 1082
Db      321 SerSerValGlnLysValPheGlnValValGlnSerThrArgProGlySylValTyr 340
QY      1083 TTAGAGAAACAAAGCTTCGATATGAGAGGCGAGCGCCCTTGCATATCCGACCTTTGCA 1142
Db      341 LeuGlyGlnThrSerSerAlaTyrGlyGlyAlaProLeuLeuSerAspThrPheAla 360
QY      1143 GCTGGCTTTATGTCGCGATTAATTTGGGCTGTGACGCGCAATGGGAATGAAAGTGTG 1202
Db      361 AlaGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaArgMetGlyIleGlnValVal 380
QY      1203 ATGAGGCAAGTATTTCTTGGAGCAGAACTACCATTTAGTGAATGAAAACTTCGATCCT 1262
Db      381 MetArgGlnValPhePheGlyAlaGlyAsnTyrHisLeuValAspGlnAsnPheAspPro 400
QY      1263 TTACCTGATTTATGCTATCTCTCTGTTCAAGAAATTTGGTGGCACCAAGTCTTATG 1322
Db      401 LeuProAspTyrTrpLeuSerLeuLeuPheLysLeuValGlyTyrTrpValLeuMet 420
QY      1323 GCAAGCGTGAAGGTTCAAGAGAGAAAGCTTCAGATATACCTTCATTCGACAAACACT 1382
Db      421 AlaSerValGlnGlySerLysArgLysLeuLysValTyrLeuHisCysThrAsnThr 440
QY      1383 GACAATCCAAAGTATTAAGAAGAGATTTAACTCTGTATGCGCATTAACCTCCATMACGTC 1442
Db      441 AspAsnProArgTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLeuHisAsnVal 460
QY      1443 ACCAAGTACTTGGCGGTAACTCTATCTTTTCTAACAAGCAAGTGGATTAATACCTTCTA 1502
Db      461 ThrLysTyrLeuAspLeuProLysProPheSerAsnLysGlnValAspLysTyrLeuLeu 480
QY      1503 AGACCTTTGGGAACTCATGATTACTTCTTCCAAATCTGCGCAACTCATGAGCTTAACCTA 1562
Db      481 ArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeuThrLeu 500
QY      1563 AAGATGTGATGATCAACCTTGCCACTTTAAATGAAAAAAGCTTCGCGCGCAGAGAGT 1622
Db      501 LysMetValAspAspGlnThrLeuProProLeuMetGlnLysProLeuLysGlySer 520
QY      1623 TCACCTGGGCTTGGCCAGCTTTCTCATATAGTTTTTTTGTGATAGAAATGCCAAGTTGCT 1682
Db      521 SerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLysValAla 540
QY      1683 GCTTGATC 1691
Db      541 AlaCysIle 543

```

```

RESULT 8
AA97635
ID AA97635 standard; protein; 543 AA.
XX
AC AA97635;
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase protein sequence.
XX
KW Heparanase; hnp1; wound healing; angiogenesis; restenosis; Scarpa;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO20010643-A2.
XX
PD 04-JAN-2001.
XX
PE 19-JUN-2000; 2000MO-IL000358.
XX
PR 25-JUN-1999; 99US-0140801P.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Pecker I, Michal I, Itzhaki H;
XX
DR WPI, 2001-137930/14.
XX
PT New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy protocols
PT for angiogenesis, restenosis, atherosclerosis, or inflammation.
XX
PS Disclosure, Page 64-65; 67pp; English.
XX
CC This sequence represents a heparanase of the invention. The heparanase
CC DNA and protein sequences are useful in wound healing, angiogenesis,
CC restenosis, atherosclerosis, inflammation, pulmonary disease,
CC neurodegenerative diseases (such as Scarpa, Alzheimer's disease, and
CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding
CC sequence is particularly useful in gene therapy
XX
SQ Sequence 543 AA:
XX
Alignment Scores:
Pred. No.: 1.2e-272 Length: 543
Score: 2842.00 Matches: 543
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.12% Indels: 0
DB: 4 Gaps: 0
US-10-676-079-3 (1-1721) x AA97635 (1-543)
QY 63 ATGCTGCTGAGCTCGAAGCTGCGCTGCGCGCGCTGATGCTGCTGCTGCGGCGG 122
Db 1 MetLeuLeuArgSerLysProAlaLeuProProProLeuMetLeuLeuLeuGlyPro 20
QY 123 CTGGGTCCTCTCTCCCTGCGCGCTGCGCGCTGCGCAAGCAAGACGTCGTGAC 182
Db 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaAspValValAsp 40
QY 183 CTGGAATTTCTTCAACCAAGAGCGCGCTGCACTGTGTAGAGCCCTCTCTGTCCTCCGAC 242
Db 41 LeuAspPhePheThrGlnGlnProLeuHisLeuValSerProSerPheLeuSerValThr 60
QY 243 ATGAGCGCAACCTGCGCAGCAGCAGCGCGGTTCTCATCTCTGAGGTTCTCCAAAGCTT 302
Db 61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProLysLeu 80

```

QY 303 CGTACCTTGGCCAGAGGCTTGTCTCTCGCTACTGAGTTGTGGCCAGACAGAC 362  
 DB 81 AGTTThleuAlaArgGlyLeuSerProAlaTyrlleuArgPheGlyGlyThrIysThrAsp 100  
 QY 363 TTCTTAATTTTGCATCCCAAGAGAAATCAACCTTTGAAGAGAGATTACTGGCAATCT 422  
 DB 101 PheIleIlePheAspProIysIysGlnSerThrPheGlnGlnArgSerTyrlProGlnSer 120  
 QY 423 CAAGTCACACAGAGATTTTGGCAATTTGATTCATCCCTCCGATGGTGGAGAGAGTTA 482  
 DB 121 GlnValAsnGlnAspIleCysIysTyrlGlySerIleProProAspValGlnGlnIlyIysLeu 140  
 QY 483 CGGTGGAATGGCCCTACCAAGAGCAATTTACTCTCCAGAACACTACCAAGAAAAATTC 542  
 DB 141 ArgLeuGlnTrpProTyrlGlnGlnGlnLeuLeuArgGlnIleTyrlGlnIlyIysPhe 160  
 QY 543 AAGAACAGACCTTACTCAAGAAAGCTGTGATGATGCTTAACACTTTTGCAAATGCTCA 602  
 DB 161 LysAsnSerThrTyrlSerArgSerSerValAspValLeuTyrlThrPheAlaAsnCysSer 180  
 QY 603 GGACTGAGTATGATCTTTGGCCCTTAAATGGGTTATTAAAGACAGCATTTGGAGTGGAC 662  
 DB 181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn 200  
 QY 663 AGTTCTAATGCTCAGTTGCTCTGCACTACTGCTCTTCCAGAGGGATTAACATTTCTTGG 722  
 DB 201 SerSerAsnAlaGlnLeuLeuLeuAspTyrlCysSerSerIlyGlyTyrlAsnIleSerTrp 220  
 QY 723 GAATGAGCAATGAACCTTAACAGTTTCTTAAAGAGCTGATATTATTTTCATCAATGGGTCG 782  
 DB 221 GlnLeuGlyAsnGlnProAsnSerPheLeuIlyIysAlaAspIlePheIleAsnGlySer 240  
 QY 783 CAGTTAAGAGAGATTAATTTCAATTGCAATTAACCTTAAAGAGCCACTTCAAAAT 842  
 DB 241 GlnLeuGlyGlnAspTyrlIleGlnLeuHisIlyLeuLeuArgIysSerThrPheIysAsn 260  
 QY 843 GCAAAAACCTATGCTCTGATGTTGGTCAAGCTCGAAGAAAGACGGCTAAGATGCTGAAG 902  
 DB 261 AlaIlyIysLeuTyrlClyProAspValGlyGlnProArgIlyThrAlaIlyIysMetLeuIys 280  
 QY 903 AGCTTCTGAGAGCGTGGTGGAGAGATGATTGATTCAGTTACATGCGATCACTACTATTGG 962  
 DB 281 SerPheLeuIlyAlaGlyIlyGlnValIleAspSerValThrTrpHisIlyTyrlIleu 300  
 QY 963 AATGACGCACTGCTCTACAGAGAAAGTTTCTTAAACCTCGATGATTAATGCAATTTTATT 1022  
 DB 301 AsnGlyArgThrAlaThrArgIlyAspPheLeuAsnProAspValLeuAspIlePheIle 320  
 QY 1023 TCATCTGTGCAAAAAGTTTTCAGGTGTGTGAGAGACAGCGCTGGCAAGAAAGTCTGG 1082  
 DB 321 SerSerValGlnIlyValIlePheGlnValIleGlnSerThrArgProGlyIlyIysValIlyTrp 340  
 QY 1083 TTAGAGAAAACAAGCTCTGCATATGAGAGCGAGCGCCCTTGCTATCCGACACTTTGCA 1142  
 DB 341 LeuGlyGlnThrSerSerAlaTyrlGlyIlyAlaProLeuLeuSerAspThrPheAla 360  
 QY 1143 GCTGCGTTATATGCGTGGATAAATGGGCTCTGACCCCGAATGGGAATTAAGTGGTG 1202  
 DB 361 AlaGlyPheMetTrpLeuAspIlyIysLeuGlyLeuSerAlaArgIleGlyIlyValIlyVal 380  
 QY 1203 ATGAGGCAAGTATCTCTTGGAGAGCAAACTACCATTAAGTGAATAAACTTCATCTCT 1262  
 DB 381 MetArgGlnValIlePhePheGlyAlaGlyIlyAsnTyrlHisIleValIlyAspIlyIysPheAspPro 400  
 QY 1263 TTACTGATTAATTTGGCTATCTCTTGTTCAGAAAATGGTGGGCAAGAGTGTAAATG 1322  
 DB 401 LeuProAspTyrlTrpLeuSerLeuLeuPheIlyIysLeuValIlyGlyThrIlyValIlyLeuMet 420  
 QY 1323 GCAAGGTCGCAAGCTTCAAGAGAGAAAGCTTCGAGTATACCTTCAATGCAACAACT 1382  
 DB 421 AlaSerValGlnIlySerIlyIysArgIlyIysLeuArgValIlyTyrlLeuHisCysThrAsnThr 440  
 QY 1383 GACAATCAAGGTATTAAGAAAGAGATTAACTCTGTAAGCCATAAACCCTCAATACGTC 1442

DB 441 AsnAsnProArgTyrlGlyGlnGlyIlyAspLeuThrIleuTyrlAlaIleAsnLeuHisAsnVal 460  
 QY 1443 ACCAAGTACTTGGCGGTATACCTATCCCTTTTCTTAAACAGCAAGTGGATTAATACCTTCTA 1502  
 DB 461 ThrIysTyrlLeuArgLeuProTyrlProPheSerAsnIysGlnValIlyAspIlyIysLeuLeu 480  
 QY 1503 AGACCTTGGAGACCTCATGATTAATCTTCCAAATCTGTCACCAATGATGCTTAATCTTA 1562  
 DB 481 ArgProLeuGlyProHisIleGlyLeuLeuSerIlySerValGlnLeuAsnGlyLeuThrIleu 500  
 QY 1563 AAGATGTGATGATCAAAACCTTGGCACTTTAAATGAAAAACCTTCGCGCCAGAAAGT 1622  
 DB 501 LysMetValIlyAspArgGlnThrIleuProLeuMetGlnIlyIysProLeuArgProGlySer 520  
 QY 1623 TCATGCGGCTTGGCCAGCTTCTCATATATGTTTGTGATTAAGAAAGCCAAAGTCTCT 1682  
 DB 521 SerLeuGlyLeuProAlaPheSerTyrlSerPhePheValIleArgAsnAlaIlyValAla 540  
 QY 1683 GCTTGCAATC 1691  
 DB 541 AlaCysIle 543  
 RESULT 9  
 ABB07813  
 ID ABB07813 standard; protein; 543 AA.  
 AC ABB07813;  
 DT 03-JUL-2002 (first entry)  
 XX  
 DE Human heparanase sequence.  
 KW Heparanase; catalytic; cytosolic; antiviral; antibacterial; enzyme;  
 anti-protocozan; neuroprotective; heparin; human.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FH Peptide 1..35 /note= "signal peptide"  
 FT Protein 36..543 /note= "mature protein"  
 FT  
 XX  
 PN US2002034810-A1.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 16-AUG-2001; 2001US-00930218.  
 XX  
 PR 20-SEP-2000; 2000US-00666390.  
 XX  
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 XX  
 PI Goldsmith O, Pecker I, Vlodevsky I, Michael I, Zcharia B;  
 DR WPI; 2002-338926/37.  
 XX  
 PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
 PT to treat various heparin-related disorders and the signal peptide is  
 PT useful in production of membrane-targeted or secreted recombinant  
 PT proteins.  
 XX  
 PS Disclosure; Fig 1a; 39pp; English.  
 CC The invention relates to an isolated avian and reptile nucleic acid,  
 CC encoding a polypeptide with heparanase catalytic activity. The signal  
 CC peptide of the nucleic acid can be used to express membrane-associated or  
 CC secreted proteins in heterologous expression systems. The encoded  
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
 CC invasion, and to intervene with pathologies associated with impaired  
 CC heparin-binding growth factors, cellular responses to heparin-binding  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,



CC cellular susceptibility to viral, protozoa and bacterial infections or  
CC disinfection of neurodegenerative plaques. The present sequence  
CC represents a human heparanase protein sequence used in similarity studies  
XX  
SQ Sequence 543 AA:

## Alignment Scores:

Pred. No.:	1,26-272	Length:	543
Score:	2842.00	Matches:	543
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	91.12%	Indels:	0
DB:	5	Gaps:	0

US-10-676-079-3 (1-1721) x ABB07813 (1-543)

```

QY 63 ATGCTGCTGCGCTGGAAGCTGCGCTGCGCGCGCTGATGCTGCTGCTGCGGCGG 122
   |||||
Db 1 MetLeuLeuArgSerIysProAlaLeuProProProLeuMetLeuLeuGlyPro 20

QY 123 CTGGGTCCTCTCTCCCTGCGCGCTGCGCGCTGCGCAAGACAGACAGCTGTGAC 182
   |||||
Db 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAlaValAsp 40

QY 183 CTGACCTTCTTCAACCCAGAGAGCCGCTGCACTGTGAGCCCTGCTCTGCTCCGTCACC 242
   |||||
Db 41 LeuAspPhePheThrGlnGlnProLeuHisLeuValSerProSerPheLeuSerValThr 60

QY 243 ATTGACGCCAAGCTGCGCGCGCGCGGTTCTCATCTCTCGGGTTCCTCAAGCTT 302
   |||||
Db 61 IleAspAlaAsnLeuAlaThrAspPheLeuLeuLeuGlySerProIysLeu 80

QY 303 CGTACCTTGGCGCAGAGGCTTGTCTCTGCGCTGACCTGAGTTGTGGACCAAGACAGAC 362
   |||||
Db 81 ArgThrLeuAlaArgGlyLeuSerProAlaIleuArgPheGlyGlyThrIleThrAsp 100

QY 363 TTCTTATTTTTCGATCCCAAGAGAGATCAACTTTGAAGAGAGATTTACTGCAATCT 422
   |||||
Db 101 PheLeuIlePheAspProIysGlySerThrPheGlnGlnArgSerIysTrpGlnSer 120

QY 423 CAAGTCAACAGAGATTTTGAATATGATTCATCTCCCTGATGTGGAGAGAGATTA 482
   |||||
Db 121 GlnValAsnGlnAspIleGlyIleGlySerIleProProAspValGlnGlnIleLeu 140

QY 483 CGGTGGAATGCGCTTACCAAGAGCAATTTGCTACTCCGAGAACCTACCAAGAAATTC 542
   |||||
Db 141 ArgLeuGlnIleProIysGlnGlnGlnLeuLeuArgGlnIleIleIleIleIlePhe 160

QY 543 AAGAACGACGCTTCTCAAGAGGCTGTGATGATGCTATACATTTTGCACAACTGCTCA 602
   |||||
Db 161 LysAsnSerThrIysSerArgSerSerValAspValLeuIleIleIleIleIleIleIle 180

QY 603 GGAAGTGAATGCTTGTGGCTTAATAGCGTTATTAAGAACAGAGATTGCAAGTGAAC 662
   |||||
Db 181 GlnLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnIlePhe 200

QY 663 AGTTCTAATGCTGATGCTCTCTGACTACTGCTCTTCAAGGGGTTAAATTTCTTGG 722
   |||||
Db 201 SerSerAsnAlaGlnLeuLeuAspIysCysSerSerIysGlyIleIleIleIleSerTrp 220

QY 723 GAATAGGCAATGAAGCTTCAAGTTTCTTAAGAGAGCTATATTTTCAATCAATGGCTCG 782
   |||||
Db 221 GlnLeuGlyAsnGlnProAsnSerPheLeuIleIleIleIleIleIleIleIleIleIle 240

QY 783 CAGTTAGAGAAAGATTAATTCATTAATGATTAATTTTAAAGAAAGTCCACTTCAAAAT 842
   |||||
Db 241 GlnLeuGlyIleAspIleIleGlnIleHisIleIleLeuLeuArgIleIleIleIleIleIle 260

QY 843 GCAAAATCTATGCTCTGATGTTGGTCAAGCTTGAAGAAAGACGAGTAAAGATGCTGAG 902
   |||||
Db 261 AlaIleLeuIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 280

QY 903 AGCTTCTGAAGGCTGTGAGAGAGATTAATTCATTAATGCTCACTACTACTATTGG 962

```

```

Db 281 SerPheLeuIleAlaGlyGlyValIleAspSerValThrThrIleIleIleIleIleIle 300
QY 963 AATGACGAGCTGCTTACCAAGAGATTTTCTAAACCTGATGATTTGACATTTTATTT 1022
   |||||
Db 301 AsnGlyArgThrAlaThrArgIleAspPheLeuAsnProAspValLeuAspIlePheIle 320

QY 1023 TCATCTGTGCAAAAGATTTTCCAGTGTGTGAGAGCAGCAGGCTGCAAGAGATCTGG 1082
   |||||
Db 321 SerSerValGlnIleValPheGlnValIleGlnIleSerThrArgProGlyIleIleValTrp 340

QY 1083 TTAGAGAAACAAAGCTTGTGATATGAGAGCGAGCGCCCTTGTATCCGACACTTTGCA 1142
   |||||
Db 341 LeuGlyGlnIleIleSerIleIleIleIleIleIleIleIleIleIleIleIleIleIle 360

QY 1143 GCTGCTTTATGCTGCTGATTAATTTGGGCTGTCAGCCGCAATGGAATAGAGTGTG 1202
   |||||
Db 361 AlaGlyPheMetThrLeuAspIleIleGlyLeuSerAlaArgMetCylIleGlnValIle 380

QY 1203 ATGAGCGAAGTATTTCTTGGAGCAGGAAATCAATTTATGAGATGAAATTCGATCCT 1262
   |||||
Db 381 MetArgGlnValPhePheGlyAlaGlnIleIleIleIleIleIleIleIleIleIleIle 400

QY 1263 TTACCTGATTAATGCTGCTATCTCTGCTTCAAGAAATGCTGGGCAAGAGGTTAAAG 1322
   |||||
Db 401 LeuProAspIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 420

QY 1323 GCAAGCGTGAAGGTTCAAGAGAGAGAGAGAGAGCTTCAGTATTCATTTGCAAAACACT 1382
   |||||
Db 421 AlaSerValGlnIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 440

QY 1383 GACAAATCAAGATTAAGAGAGAGATTTACTGTGATGCAATTAACCTCAATACGTC 1442
   |||||
Db 441 AspAsnProArgIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 460

QY 1443 ACCAAGTACTGGGCTTACCTCTATCTTTTCTAAACAGTGAATTAATCTTCTTA 1502
   |||||
Db 461 ThrIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 480

QY 1503 AGACCTTGGGACCTCATGATGATTAATTTTCAATCTGTCAATCAATGATGCTTAATCTTA 1562
   |||||
Db 481 ArgProLeuGlyProIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 500

QY 1563 AGATGCTGATGATTAACCTTGGCACTTTTATGAGAAACCTTCCGCGCAGAGAGT 1622
   |||||
Db 501 LysMetValAspAspGlnThrLeuProProLeuMetGlnIleIleIleIleIleIleIleIle 520

QY 1623 TCATGCGGCTGCGCAGCTTCTCATATAGTTTGTGATAGAAATGCAAGTGTGCT 1682
   |||||
Db 521 SerLeuGlyLeuProAlaPheSerIysPhePheValIleIleIleIleIleIleIleIleIle 540

QY 1683 GCTTGCATC 1691
   |||||
Db 541 AlaCysIle 543

RESULT 10
AAV17082
ID AAV17082 standard; protein; 543 AA.
XX
AC AAV17082;
XX
DT 21-JUL-1999 (first entry)
XX
DE Human heparanase enzyme.
XX
KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW metachasis; angiogenesis; wound healing; angioplasy-induced revascularis;
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW human; HSPG.
XX
OS Homo sapiens.
XX
PN MO9921975-A1.

```

XX 06-MAY-1999.  
 PD  
 XX 28-OCT-1998; 98MO-AU000898.  
 PF  
 XX 28-OCT-1997; 97AU-0000062.  
 PR 09-DEC-1997; 97AU-00000812.  
 XX  
 XX (AUSU ) UNIV AUSTRALIAN NAT.  
 PA  
 PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;  
 XX WPI, 1999-312956/26.  
 DR N-PSDB; AAX37259.  
 XX  
 XX Polynucleotides encoding mammalian endoglucuronidases, especially  
 PT heparanases, useful to promote wound healing.  
 CC  
 PS Claim 6; Page 69-73; 112pp; English.

The invention relates to nucleic acid sequences that encode heparanase enzymes having endoglucuronidase activity. Recombinant heparanases are capable of removing the HS side chain from heparan sulfate proteoglycan (HSPG). Sulfated oligosaccharides, sulphates or HSPG can be used to inhibit heparanase, this is useful for treatment of a physiological or medical condition associated with elevated heparanase activity, such as metastasis, angiogenesis, wound healing, angioplasty-induced restenosis, arteriosclerosis, atherosclerosis and inflammation. The human, murine and rat heparanases can be used to enhance wound healing, especially CC associated with tissue development and repair. The conditions mentioned above can be diagnosed using specific antibodies, and also using primers CC and probes specific for the heparanase polynucleotides. Other uses of the CC heparanases include sequencing sulfated molecules such as HSPG. The CC present sequence represents a human heparanase

XX Sequence 543 AA;

XX Alignment Scores:

Pred. No.: 3 01e-272 Length: 543  
 Score: 2838.00 Matches: 542  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.82% Mismatches: 0  
 Query Match: 90.99% Indels: 0  
 DB: 2 Gaps: 0

US-10-676-079-3 (1-1721) x AAY17082 (1-543)

QY 63 ATGCTGCTGCGCTCGAAGCGCTGCGCGCGCGCTGATGCTGCTGCGGCGCG 122  
 DB 1 MetLeuLeuArgSerIleProAlaLeuProProLeuMetLeuLeuLeuGlyPro 20  
 QY 123 CTGGGTCCTCTCCCTGCGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCGCG 182  
 DB 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaValValaIle 40  
 QY 183 CTGGAAGCTTTCACCGAGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCG 242  
 DB 41 LeuAspPhePheThrGlnProLeuHisLeuValSerProSerPheLeuSerValThr 60  
 QY 243 ATGAGCGCAACCTGCGCGCGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCG 302  
 DB 61 IleAspAlaHisLeuAlaThrAspProAlaGlyPheLeuLeuLeuGlySerProLeu 80  
 QY 303 CGTACCTTGGCGAGAGCTTGTCTCTGCGTACCTGAGGTTGTGTGCGCGCG 362  
 DB 81 ArgThrLeuAlaArgIleLeuSerProAlaIleLeuArgPheGlyGlyThrIleThrAsp 100  
 QY 363 TTCCTAATTTGGATCCCAAGAGGAATCACTTTGAGAGAGAGAGTACTGGCAAT 422  
 DB 101 PheLeuIlePheHisProLeuGlySerGlnSerThrPheGlnGlnArgSerIleTrpGlnSer 120  
 QY 423 CAGTCACACAGGATATTTGCAATATGATCATCTCTGATGTGAGAGAGATTA 482

DB 121 GlnValAsnGlnAspIleCysIleYrGlySerIleProProAspValGlnGlnLeu 140  
 QY CGGTGGAAATGGCCCTACACAGAGCAATGCTACTCCGAGAACTACAGAAAAAGTTC 542  
 DB 141 ArgLeuGlnTrpProIleYrGlnGlnLeuLeuLeuArgGlnHisIleYrGlnIleAspPhe 160  
 QY 543 AAGAACAGACCTACTCAAGAGCTGTAGATGTGCTATACACTTTTGGCAACTGCTCA 602  
 DB 161 LysAsnSerThrIleYrSerArgSerSerValAspValLeuIleYrThrPheAlaHisCysSer 180  
 QY 603 GCACTGCACTTGATCTTTGGCTTAAATGCGTTATTAAGACAGCAAGATTTGCAATG 662  
 DB 181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn 200  
 QY 663 AGTTCTAATGCTACGTTGCTCTGACACTGCTCTTCAAGGGGATTAACATTTCTGG 722  
 DB 201 SerSerAsnAlaGlnLeuLeuLeuAspIleYrCysSerSerIleGlyTrpAsnIleSerTrp 220  
 QY 723 GAACCTAGGCAATGAACTTACAGATTTCTTAAAGAGCTGATATTTTCATCAATGGCTG 782  
 DB 221 GlnLeuGlnAsnGlnProAsnSerPheLeuIleYrValAspIlePheIleAsnGlySer 240  
 QY 783 CAGTTAGAGAAATTAATTAATTCATTAATGCACTTCAAGAAAGTCACTTCAAAAT 842  
 DB 241 GlnLeuGlyGlnAspPheIleGlnLeuHisIleYrLeuLeuArgIleSerThrPheIleAsn 260  
 QY 843 GCAAAACCTTATGCTGCTGATGTTGTCAGCTGAGAGAAAGAGCGGCTAAGATGCTAG 902  
 DB 261 AlaIleLeuIleYrGlyProAspValGlyGlnProArgGlySerAlaIleLeuLeuIle 280  
 QY 903 AGCTTCTGAAAGGCTGTGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 962  
 DB 281 SerPheLeuIleValIleGlyGlyGlyValIleAspSerValThrPheHisIleYrIleLeu 300  
 QY 963 AATGAGCGACTGCTACCGAGAGATTTGTAACCTGATGATGATGATGATGATGAT 1022  
 DB 301 AsnGlyArgThrAlaThrArgIleAspPheLeuAsnProAspValLeuAspIlePheIle 320  
 QY 1023 TCATCTGCGCAAAAAGTTTTCAGGTGCTGAGAGACACAGCGCTGCGCAAGGCTGCG 1082  
 DB 321 SerSerValGlnIleValPheGlnAlaValGlnSerThrArgProGlyIleValValIleTrp 340  
 QY 1083 TTAGAGAGAAAGAGCTGCTGATGATGAGAGCGCGCGCTGCTGATCGAGACCTTTGCA 1142  
 DB 341 LeuGlyGlnThrSerSerIleIleYrGlyGlyAlaPheLeuLeuSerAspThrPheAla 360  
 QY 1143 GCTGCTTATGCTGCTGATTAATTTGGGCTGTCAAGCGCAATGGAATTAAGAGTGTG 1202  
 DB 361 AlaGlyPheMetTrpLeuAspIleLeuGlyLeuSerIleAlaArgMetGlyIleGlnValVal 380  
 QY 1203 ATGAGGCAAGATTTCTTTGAGAGAGAGAACTAACATTTAGTGAGTAAGTCTGATG 1262  
 DB 381 MetArgGlnValPhePheGlyAlaGlyAsnIleValIleValIleAspGlnAspPheAsp 400  
 QY 1263 TTACCTGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1322  
 DB 401 LeuProAspIleYrTrpLeuSerIleLeuPheLeuIleValGlyThrIleValLeuMet 420  
 QY 1323 GCAAGCGTGAAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382  
 DB 421 AlaSerValGlnGlySerIleValArgIleValArgValIleLeuHisCysThrAsnThr 440  
 QY 1383 GACATTCGAAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1442  
 DB 441 AspAsnProArgIleYrGlyGlyIleAspLeuThrLeuIleValIleAsnLeuHisVal 460  
 QY 1443 ACCAAGTCTTGGGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1502  
 DB 461 ThrIleYrIleValGlnLeuProIleYrProPheSerAsnIleGlnValAspIleYrIleLeu 480  
 QY 1503 AGACCTTTGGAGCTCATGATTAATTTCAAAATGCTGCAAGTCAATGCTGCTAACTTA 1562  
 DB 481 ArgProLeuGlyProHisGlyLeuLeuSerIleSerValGlnLeuAsnGlyLeuThrIleu 500

QY 1563 AAGATGGTGAATGATCAAACTTGGCCACTTTAATGAAAAAAGCTTCGCCGAGAACT 1622  
DB 501 LysMetValaAspArgInThrLeuProLeuMetGluLysProLeuArgProGlySer 520  
QY 1623 TCACGTGGCTGGCCAGCTTTCTCATATAGTATTTTGGATAGAAAAATGCCAAAGTTGCT 1682  
DB 521 SerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLysValAla 540  
QY 1683 GCTTGCAATC 1691  
DB 541 AlaCysIle 543  
RESULT 11  
AAB86206  
ID AAB86206 standard; protein, 543 AA.  
AC AAB86206;  
XX 24-AUG-2001 (first entry)  
DT  
XX Human heparanase inhibitor protein.  
DE  
XX Heparanase; inhibitor; cardiac insufficiency; cardiast; nephrotropic;  
KM hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;  
KM primary cardiomyopathy; peripheral edema; pulmonary congestion;  
KM hepatic congestion; hydrothorax; ascites; nocturia; human.  
XX Homo sapiens.  
OS  
XX DE19955803-AI.  
PN 23-MAY-2001.  
XX 19-NOV-1999; 99DE-01055803.  
PF 19-NOV-1999; 99DE-01055803.  
XX 19-NOV-1999; 99DE-01055803.  
PR  
XX (KNOL ) KNOLL AG.  
PA  
PI Herr D, Hahn A, Laux V;  
XX WPI: 2001-368371/39.  
DR N-PSDB; AAH20940.  
XX  
PT Treatment or prevention of cardiac insufficiency and related conditions,  
PT e.g. pulmonary congestion and dyspnoea, comprises administration of  
PT heparanase inhibitor.  
XX  
PS Disclosure; Page 11-13; 16pp; German.  
XX  
XX This invention describes a novel heparanase inhibitor which can be used  
XX for the treatment or prevention of cardiac insufficiency and associated  
XX indications, symptoms and/or malfunctions. The heparanase inhibitor of  
XX the invention has cardiant, nephrotropic and hepatotropic activity. The  
XX products of the invention can be used in human and veterinary medicine,  
XX for the treatment or prevention of congestive heart failure e.g. primary  
XX cardiomyopathy. Associated conditions treated or prevented with the  
XX inhibitor are especially peripheral edemas, pulmonary and hepatic  
XX congestion, dyspnoea, hydrothorax and ascites. Renal problems e.g.  
XX nocturia can also be treated. This sequence represents the human  
XX heparanase protein described in the method of the invention  
XX  
SQ Sequence 543 AA;  
Alignment Scores:  
Pred. No.: 3 01e-272 Length: 543  
Score: 2838.00 Matches: 542  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.82% Mismatches: 0  
Query Match: 90.99% Indels: 0  
DB: 4 Gaps: 0

US-10-676-079-3 (1-1721) x AAB86206 (1-543)  
QY 63 ATCTGTGGCTCGAAGCCTGGCTGCCGCCCGCTGATGCTGCTGCTGCGGCGG 122  
DB 1 MetLeuLeuArgSerIysProAlaLeuProProProLeuMetLeuLeuLeuGlyPro 20  
QY 123 CTGGGTCCTCTCTCCCTGGCGCCCTGCGCCGACCTGGCCAGACAGACGTCGTGAC 182  
DB 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAspValAlaAsp 40  
QY 183 CTGACCTCTTCCACCCAGAGCCGCTGACCTGTGAGCCCTGCTGCTGCTGCTGCTG 242  
DB 41 LeuAspPhePheThrGlnGlnProLeuHISLeuValSerProSerPheLeuSerValThr 60  
QY 243 ATTGAGCCCAAGCTGGCCAGACCCGCGGTTCTCATCTCTGCGTTCTCCAAAGCTT 302  
DB 61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuLysProLysLeu 80  
QY 303 CGTACCTGGCCAGAGGCTTGTCTCTGCGTACCTGAGGTTGTGGCACCAAGACAGAC 362  
DB 81 ArgThrLeuAlaArgGlyLeuSerProAlaTyrLeuArgPheGlyGlyThrLysThrAsp 100  
QY 363 TTCCTAATTTTGGATCCCAAGAGAAATCAACCTTTGAAGAGAGATTAGGCAATCT 422  
DB 101 PheLeuIlePheAspProLysGlySerThrPheGlnGlnArgSerTyrTrpGlnSer 120  
QY 423 CAAGTCACACGAGATATTTGCAATATGATTCATCTCTGATGTGAGAGAGAAATTA 482  
DB 121 GlnValAsnGlnAspIleCysLysTyrGlySerIleProProAspValGlnGlnLysLeu 140  
QY 483 CGGTGGATGGCCCTTACCAAGAGCAATGCTNACTCCGAGAACATCAACGAAAAAGTTC 542  
DB 141 ArgLeuGlnTrpProLysGlnGlnGlnLeuLeuArgGlnIleTyrGlnLysLysPhe 160  
QY 543 AAGACAGCAGCCTACTCAAGAAAGCTGTAGATGTGCTATACCTTTGCAAACTGCTCA 602  
DB 161 LysAsnSerThrTyrSerIysSerValAspValLeuTyrThrPheAlaAsnCysSer 180  
QY 603 GGACTGACTGATCTTGGCCCTAAATGCGTTATTAAGAACAGACAGATTGGAGTGAAC 662  
DB 181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn 200  
QY 663 AGTTCTAATGCTCAGTTGCTCTGACATCTGCTCTTCCAGGGGATTAACATTTCTTGG 722  
DB 201 SerSerAsnAlaGlnLeuLeuLeuAspTyrCysSerSerLysGlyTyrAsnIleSerTrp 220  
QY 723 GAATGAGCAATGAACCTTAACAGTTTCTTAAGAGGCTGATTTTCAATGAGGCTG 782  
DB 221 GlnLeuGlyAsnGlnProAsnSerPheLeuLysLysAlaAspIlePheIleAsnGlySer 240  
QY 783 CAGTTAGAGAAAGATTATTAATTCATTGATTAATTCATTAAGAAAGTCCACTTCAAAAT 842  
DB 241 GlnLeuGlyGlnAspPheIleGlnLeuHISLysLeuLeuArgLysSerThrPheLysAsn 260  
QY 843 GCAAAACCTTAGTGCTGATGTTGTGCTGACCTCCGAAAGAAAGCGCTTAAGTCTGAAG 902  
DB 261 AlaLysLeuTyrGlyProAspValGlyGlnProArgArgLysThrAlaLysMetLeuLys 280  
QY 903 AGCTTCCTGAAGCGTGTGAGAGAGATTTGATTCATGATTCATGATTCATGATTCAT 962  
DB 281 SerPheLeuLysAlaGlyLysGlnValIleAspSerValThrTrpHisTyrTyrLeu 300  
QY 963 AATGACGAGCTGCTACACGAGAGATTTTCTTAACCTGATGTATTTGACATTTTAT 1022  
DB 301 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320  
QY 1023 TCATCTGTGCAAAAAGTTTCCAGGTGTTGAGAGACACAGGCTGGCAAGAGCTTGG 1082  
DB 321 SerSerValGlnLysValPheGlnValValGlnSerThrArgProGlyLysLysValTrp 340  
QY 1083 TTAGAGAAACAAGCTGCTCATATGAGGCGGAGGCGCTTGTCTATTCGACACTTTGGA 1142  
DB 341 LeuGlyGlnThrSerSerAlaTyrGlyGlyAlaProLeuLeuSerAspThrPheAla 360

QY	1143	GCTGGCTTTATGTGGCTGGATAAATATGGGCTGTGCAGCCCAATGGGAATGGAAGTGTG	1202
Db	361	AlaGlyPheMetCrrPLeuAspLysLeuGlyLeuSerAlaArgMetGlyIleGluVal	380
QY	1203	ATGAGCGCAAGTATCTTTGGAGCAGAGAAACTACCATTTAGTGATGAAACTTCGATCCT	1262
Db	381	MetArgGlnValPhePheGlyAlaGlyAsnTyrHisLeuValAspGluAsnPheArgPro	400
QY	1263	TTACTGATTTATTTGGCTATCTCTCTGTTCAGAAATTTGGTGGGACCAAGTGTATG	1322
Db	401	LeuProAspTyrTrrPLeuSerLeuLeuPheLeuValGlyThrLysValLeuMet	420
QY	1323	GCAAGCGCGCAAGGTTCCAAAGAGAAAGAGCTCGAGTATACCTTCATTTGCACAAACACT	1382
Db	421	AlaSerValGlnGlySerLysArgThrGlySerLeuAlaGlyThrLysCysThrAsnThr	440
QY	1383	GACAAATCCAAAGTATTAAGAAGAGATTAACTGTATGGCATTAACCTTCATTAACGTC	1442
Db	441	AspAsnProArgTyrLysGluGlyAspLeuThrLeuTyrAlaIleAsnLeuHisAsnVal	460
QY	1443	ACCAAGTACTTGGCGGTTACCTTATCCCTTTTCTAACAAAGCAAGTGATTAATACCTTTCTA	1502
Db	461	ThrLysTyrLeuAspGlyLeuProTyrProPheSerAsnLysGlnValAspLysTyrLeuLeu	480
QY	1503	AGACCTTTGGGAGCCCTCATGATTTACTTTCCAAATCTGTGCAACTCAATGAGTCTAACTTA	1562
Db	481	ArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAspGlyLeuThrLeu	500
QY	1563	AAGATGGTGATGATGATCAAACTTCGCACTTTAATGAAAAACCTTCCGGCCAGAGAAGT	1622
Db	501	LysMetValAspAspArgInThrLeuProLeuMetGluLysProLeuArgProGlySer	520
QY	1623	TCATCTGGGCTTGGCCAGCTTCTCATATATGTTTTTTTGATTAAGAAATGCCAAAGTTGCT	1682
Db	521	SerLeuGlyLeuProAlaPheSerTyrSerPheValIleArgAsnAlaLysValAla	540
QY	1683	GCTTGCAATC 1691	
Db	541	AlaCysIle 543	
RESULT 12			
ADD18950	ID	ADD18950 standard; protein; 543 AA.	
XX	AC	ADD18950;	
XX	DT	15-JAN-2004 (first entry)	
XX	DE	Human disease related protein SeqID439.	
KW	human, disease state; cytosolic; antiinflammatory; ophthalmological;		
KW	antiatherosclerotic; vulnerary; gene therapy;		
KW	hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;		
KW	inflammation; erythropoiesis; glycolysis; gluconeogenesis;		
KW	glucose transportation; catecholamine synthesis; iron transport;		
KW	nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;		
KW	retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;		
KW	inflammatory condition; wound healing.		
XX	OS	Homo sapiens.	
XX	PN	W02003018621-A2.	
PD	XX	06-MAR-2003.	
XX	PF	23-AUG-2002; 2002WO-GB003892.	
XX	PR	23-AUG-2001; 2001GB-00020558.	
XX	PR	05-OCT-2001; 2001GB-00024037.	
XX	XX	(OXFO-) OXFORD BIOMEDICA UK LTD.	

PI	Kingman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;
DR	WPI: 2003-290046/28.
DR	N-PSDB; ADD18951.
XX	
XX	New substantially purified polypeptide, useful for diagnosing or treating
PT	a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
PT	injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
PT	wound healing.
XX	
PS	Claim 25; SEQ ID NO 439; 424pp; English.
CC	
CC	This invention relates to novel human genes and gene product which are
CC	implicated in certain disease states. Compounds which modulate the
CC	proteins of the invention may have cytostatic, antiinflammatory,
CC	ophthalmological, antiarteriosclerotic or vulnary activities. The
CC	sequences of the invention may be useful for gene therapy. The invention
CC	may be useful for diagnosing or treating a hypoxia-regulated condition,
CC	such as tumorigenesis, angiogenesis, apoptosis, inflammation,
CC	erythropoiesis, or the biological response to hypoxia conditions
CC	including processes such as glycolysis, gluconeogenesis, glucose
CC	transportation, catecholamine synthesis, iron transport or nitric oxide
CC	synthesis. The disease includes cancer, ischemic conditions, reperfusion
CC	injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
CC	inflammatory conditions or wound healing. The present sequence is that of
CC	a disease related protein of the invention.
XX	
SQ	Sequence 543 AA;
Alignment Scores:	
Pred. No.:	3,01e-272
Score:	2838.00
Percent Similarity:	100.00%
Best Local Similarity:	99.82%
Query Match:	90.99%
DB:	7
	Gaps: 0
US-10-676-079-3 (1-1721) x ADD18950 (1-543)	
QY	ATGCTGCTGCGTCCGAAGCTGCGTGGCGCGCGCGTGAATGCTGCTCTGCGGCG 122
Db	1 MelLeuLeuAArgSerIleProAlaLeuProProIleuMetLeuLeuLeuGlyPro 20
QY	123 CTGGGTCCCTCTCCCTGCGCGCTGCGCCGACCTGGCGGCAAGCAAGCACTGTGGAC 182
Db	21 LeuGlyProLeuSerProGlyValAlaLeuProArgProAlaGlnAlaGlnValAlaSp 40
QY	183 CTGGACTTCTTACCCAGGAGCGGTGACCTGCTGAGCCCTGTCTCTGTCCTGACC 242
Db	41 LeuAspPhePheThrGlnGluProLeuHisIleuValSerProSerPheLeuSerValThr 60
QY	243 ATTGAAGCCCAACCTGGCCACGGAGCCGGGGTTCCTCATCTCCCTGGGTTCTCCAAAGCTT 302
Db	61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleuLeuGlySerProLysLeu 80
QY	303 CGTACCTTGCCAGAGGCTTGTCTCTCTGCTGACCTGAGGTTTGGTGAGCAACAGAGAC 362
Db	81 ArgThrIleuAlaArgGlyLeuSerProAlaIleuArgPheGlyGlyThrIlySerThrAsp 100
QY	363 TTCCTAATTTTGCATCCCAAGAGGAATCAACTTTGAAGAGAGATTACTGGCAATCT 422
Db	101 PheIleuIlePheAspProLysIleGluSerThrPheGlnGluArgSerTyrTrpGlnSer 120
QY	423 CAAGTCACACAGATATTTTGGAAATATGATCCATCCCTCTGATGTGGAGAGGAAGTTA 482
Db	121 GlnValAsnGlnAspIleCysLysTyrIleSerIleProProAspValGlnGluLysLeu 140
QY	483 CGGTGGAAATGAGCCCTACACAGAGGCAATGTCTACTCCAGAAACCTACAGAAAAAGTTC 542
Db	141 ArgLeuGluTrpProTyrGlnGlnGlnIleuLeuLeuArgGlnHisTyrGlnIlyLysPhe 160
QY	543 AAGAAAGCACTACTCAAGAAAGCTGTGATGATGTCTATACACTTTTGGCAAATGCTCA 602

```

Db      161  LysansSerThrTySerArgSerSerValAspValLeuTyThrPheAlaAsnCySer 180
Qy      603  GGAAGTGAATCTTGGCTTAATAGCTTATTAAAGACAGAGATTTGACGTGAGAC 662
Db      181  GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTyrAsn 200
Qy      663  AGTTCTAATGCTCAGTTGCTCTGAGACTGCTCTTCCAAAGGGGATTAACATTTCTTGG 722
Db      201  SerSerAsnAlaGlnLeuLeuAspTyrcysSerSerTyrcysGlyTyrAsnIleSerTyr 220
Qy      723  GAAGTGAAGTGAATCAAGTTTCTTAAGAGGCTGATTAATTCATCAATAGGGCTG 782
Db      221  GlnLeuGlyAsnGlnProAsnSerPheLeuLeuValAspIlePheIleAsnGlySer 240
Qy      783  CAGTTAGAGAAAGATTAATTCATTAATTCATTAATTCATTAATTCATTAATTCATTAAT 842
Db      241  GlnLeuGlyLeuAspPheIleGlnLeuHibysLeuLeuArgTyrSerThrPheLeuAsn 260
Qy      843  GCAAACTCTATGCTCTGATGTTGTCAGCTCGAAGAAAGACGGCTTAAGATGCTGAG 902
Db      261  AlAlbLeuLeuTyrcysProAspValGlyGlnProArgTyrGlyThrAlaIlysmetLeuLys 280
Qy      903  AGCTTCCTGAAGGCTGGAGAGAGATTAATTCATTAATTCATTAATTCATTAATTCATTAAT 962
Db      281  SerPheLeuLeuValAlaGlyGlyGlyValIleAspSerValThrPheIleAsnGlySer 300
Qy      963  AATGACGAGCTGCTTACAGGAGAGATTTCTAAACCTGATGATTAATTCATTAATTCATTAAT 1022
Db      301  AsnGlyTyrGlnThrAlaThrArgGlyAspPheLeuLeuProAspValLeuAspIlePheIle 320
Qy      1023  TCATCTGTGCAAAAGTTTCCAGGTGTTGAGAGACACAGGCTGGCAAGAGTCTGG 1082
Db      321  SerSerValGlnTyrcysValPheGlnValIleGlnSerThrArgProGlyLysValTyr 340
Qy      1083  TTAGAGAAACAAGCTCTGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1142
Db      341  LeuGlyGlnThrSerSerAlaTyrGlyGlyAlaProLeuLeuSerAspThrPheAla 360
Qy      1143  GCTGGCTTATATGTCGTGATTAATGGGCTGTACGCCGATGGAGATTAAGAGTGTG 1202
Db      361  AlAlbTyrcysMetTyrLeuAspTyrcysLeuGlyLeuSerAlaArgMetGlyIleGlyVal 380
Qy      1203  ATGAGGCAATATCTTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1262
Db      381  MetArgGlnValPhePheGlyAlaGlyAlaAsnTyrcysLeuValAspGlyLeuAspPro 400
Qy      1263  TTACCTATATATGTCGTATCTCTGTCAGAAATTTGGGGGACCAAGGTTTATG 1322
Db      401  LeuProAspTyrTyrLeuSerLeuPheLeuValGlyThrLysValLeuMet 420
Qy      1323  GCAAGCTGAGAGGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382
Db      421  AlaSerValGlnGlySerLysArgTyrGlyLeuArgValTyrLeuHibysTyrAsnThr 440
Qy      1383  GACAAATCAAGGTTTAAGAGAGAGATTAATCTGATGTCATTAACCTTCATTAAGCTC 1442
Db      441  AspAsnProArgTyrLysGlnGlyAspLeuThrLeuTyrcysLeuHibysAsnVal 460
Qy      1443  ACCAAGTCTGCGGTACCTATCTCTTCTTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1502
Db      461  ThrLysTyrLeuAspGlyLeuProTyrProPheSerAsnGlyGlnValAspTyrTyrLeu 480
Qy      1503  AGACCTTTGGAGCTCATGATTAATCTTCCAAATCTGTCCAACTCAATGCTTAAGCTTA 1562
Db      481  ArgProLeuGlyProHibysGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeuThrLeu 500
Qy      1563  AAGATGTTGATGATCAAACTTGGCACTTTAATGAAAAAAGCTCTCGGCGCAGAGAGT 1622
Db      501  LysMetValAspAspGlnThrLeuProLeuMetGlnLysProLeuAspProGlySer 520
Qy      1623  TCACGGGCTTGGCAGGCTTTCTCATATAGTTTGTGATTAAGAAATGCAAGTGTGCT 1682
Db      521  SerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaIlyValAla 540

```

```

Qy      1683  GCTTGCATC 1691
Db      541  AlaCysIle 543

RESULT 13
ID AAB88361 standard; protein; 543 AA.
XX
XX AAB88361;
AC
XX
XX 23-MAY-2001 (first entry)
DT
XX
XX Human membrane or secretory protein clone PSEC0090.
DE
XX
XX Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes.
XX
XX Homo sapiens.
OS
XX
XX EP1067182-A2.
PN
XX
XX 10-JAN-2001.
PD
XX
XX 07-JUL-2000; 2000EP-00114090.
PF
XX
XX 08-JUL-1999; 99JP-00194179.
PR
XX 11-JAN-2000; 2000JP-00118775.
PR 02-MAY-2000; 2000JP-00183766.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Oka T, Isogai T, Mishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI
XX WPI; 2001-093989/11.
DR
XX N-PDB; AAF93788.
XX
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development.
XX
XX Claim 1; SEQ ID NO 90; 609pp + Sequence listing; English.
XX
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by AAB88317
CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
CC invention. The invention also includes methods for the production of
CC antibodies directed against the proteins, and cDNA sequences, which can
CC be used in vaccines. The polynucleotide sequences can be used in gene
CC therapy. The polynucleotide sequences and the proteins they encode may be
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate secretory protein/membrane protein expression. The
CC nucleic acids and complementary sequences may also be used as DNA probes
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
CC and quantitate the presence of similar nucleic acid sequences in samples.
CC They may also be used to study the expression and function of secretory
CC proteins/membrane polypeptides and their role in metabolism. The
CC polypeptides may be used as antigens in the production of antibodies
CC against them and in assays to identify modulators (agonists and
CC antagonists) of expression and activity. The antibodies and antagonists
CC may also be used as therapeutic agents to down regulate expression and
CC activity. The antibodies may also be used as diagnostic agents for
CC detecting the presence of the polypeptides in samples (e.g. by enzyme
CC linked immunosorbent assay (ELISA)). Examples of diseases which may be
CC treated include rheumatoid arthritis and diabetes
XX
SQ Sequence 543 AA;

```

Alignment Scores: 4.68e-271 Length: 543  
Pred. No.: 2826.00 Matches: 540  
Score: 99.82% Conservative: 2  
Percent Similarity: 99.45% Mismatches: 1  
Best Local Similarity:

Query Match: 90.61% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-676-079-3 (1-1721) x AAB88361 (1-543)

QY 63 ATGTCGTGCGCTCGAAGCCTGCGTCCGCGCGCTGATGCTGCTGCGGCGCG 122  
 DB MetLeuLeuArgSerIleProAlaLeuProProLeuMetLeuLeuLeuGlyPro 20  
 QY 123 CTGGGTCCCTCTCCCTGCGCGCTGCGCGCGCTGCGCGAAGCAGAGAGCTGTCAG 182  
 DB LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAspValAlaSer 40  
 QY 183 CTGACCTTCCTTACCCAGAGCGCGCTGACCTGAGAGCGCTGCTGCTGCTGCGCGAC 242  
 DB LeuAspPhePheThrGlnIleProLeuHisLeuValSerProSerPheLeuSerValThr 60  
 QY 243 ATTGACGCCAACCTGCGCGCGAGCGCGCTTCTCATCTCTGCGGTCTTCCAAAGCTT 302  
 DB IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProIlySer 80  
 QY 303 CGTACCTTGGCGCAGAGGCTTGTCTCTGCGTACCTGAGGTTGGTGGCACCAAGACAGAC 362  
 DB ArgThrLeuAlaArgGlyLeuSerProAlaTyIleuArgPheGlyGlyThrIlySerThrAsp 100  
 QY 363 TTCCTAATTTTGCATCCCAAGAGGAATGACCTTTGAAGAGAGAGTACTGGCAATCT 422  
 DB PheLeuIlePheAspProIlySerIlySerGlnSerThrPheGlnIleArgSerTyIrrGlnSer 120  
 QY 423 CAAGTCAACAGAGATATTGCAATATGATGCATCCCTCTGATGTGGAGAGAGATTA 482  
 DB GlnValAlaAsnGlnAspIleCysIlySerTyIrrIlySerIleProProAspValGlnIlyIlySer 140  
 QY 483 CCGTGGGAATGGCCCTTACCAAGAGCAATTCCTACCTCCAGAAACCTACCAAGAAAAGTTCT 542  
 DB ArgLeuGlnIlyrrProIlyrrGlnIlyGlnIleuLeuLeuArgGlnHisIlyrrGlnIlyIlySerPhe 160  
 QY 543 AAGAACAGACCTTACTCAAGAGAGCTGTGATGATGCTATACACTTTTGAAGAGCTCA 602  
 DB LysAsnSerThrIlySerArgSerSerValAspValLeuTyrrThrPheHisAsnIlySerSer 180  
 QY 603 GGACTGACCTTGATCTTTGGCTTAATGCTTAATGCTTAATGAACAGAGATTTGACGTGAAC 662  
 DB GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuGlyThrAlaAspLeuGlnIrrPheAsn 200  
 QY 663 AGTTCTAATGCTCAGTTGCTCTGACATACCTGCTTCCAGAGGGGTATAACATTTCTTGG 722  
 DB SerSerAsnAlaGlnIleuLeuLeuAspTyrrCysSerSerIlyGlyTyrrAsnIlySerIrr 220  
 QY 723 GAACTAGGCAATGAACCTTAAGAGCTTCCCTTAAGAGGGCTGATATTTCATCAATGGGTG 782  
 DB GlnLeuGlyAsnGlnIrrProAsnSerPheLeuIlyIlyValAspIlePheIleAsnGlySer 240  
 QY 783 CAGTTAGAGAGATTAATTAATTCATTAATTCATTAATTCATTAAGAGAGTCCACTTCAAAAAT 842  
 DB GlnLeuGlyGlnAspPheIleGlnIleuHisIlyIlyLeuArgIlySerThrIlyIlyAsn 260  
 QY 843 GCAAAAACCTTAAGCTCTGATGTTGCTCAGCTCCAGAAAGACGGCTTAAGTCTGAAG 902  
 DB AlaIlyIlyLeuTyrrGlyProAspValGlyGlnProArgIlySerThrAlaIlyIlyMetLeuIly 280  
 QY 903 AGCTTCCTGAAGGCTGGTGAAGAGATTTGATTCATTCATTCATTCATTCATTCATTCATTCAT 962  
 DB SerPheLeuIlyAlaGlyIlyGlnIlyValIleAspSerValThrIrrHisIlyTyrrIlyLeu 300  
 QY 963 AATGACGAGACTGCTACAGAGAGATTTTCTAAACCTGATGATTTGACATTTTAT 1022  
 DB AangIlyArgThrAlaThrArgIlyAspPheLeuAsnProAspValLeuAspIlePheIle 320  
 QY 1023 TCATCTGTGCAAAAAGTTTCCAGGTGGTGGTGAAGACCAAGGCTGGCAAGAGGCTCTGG 1082  
 DB SerSerValGlnIlyValPheGlnIlyValIlyGlnSerThrArgProGlyIlyIlyValIlyrr 340

QY 1083 TTAGAGAAACAAGCTGTGATATGAGAGCGAGCGCTTGTATCCGACACTTTTGA 1142  
 DB LeuGlyGlnIrrThrSerSerAlaHisGlyIlyGlyAlaAlaProLeuLeuSerAspThrPheAla 360  
 QY 1143 GCTGGCTTATATGTGGCTGATTAATTTGGGCTGTTCAGCCGCAATGGAGATTAAGAGTGTG 1202  
 DB AlaGlyPheMetIrrPheAspIlyLeuGlyIlyLeuSerAlaArgMetGlyIleGlnIlyVal 380  
 QY 1203 ATGAGGCAAGATATCTTTTGGAGCGAGAACTTAACATTTAAGTGAATGAACCTGATCT 1262  
 DB MetArgGlnValPhePheGlyAlaGlyAsnTyrrHisLeuValAspGlnAsnPheAspPro 400  
 QY 1263 TTACCTGATTAATTTGGCTATCTCTCTGTTCAAGAAATGGTGGGCAACCAAGTGTATG 1322  
 DB LeuProAspTyrrIrrPheLeuSerIleuLeuPheIlyIlyLeuValGlyThrIlyValLeuMet 420  
 QY 1323 GCAAGCGTGAAGGTTTCAAGAGAGAGAGCTTCGAGTATACCTTCAATGCAACAACACT 1382  
 DB AlaSerValGlnIlySerIlyArgIlyAspIlyValTyrrLeuHisCysThrAsnThr 440  
 QY 1383 GACATCCAGAGTATTAAGAGAGAGATTTAATCTGTATGCGCATTAACCTCAATAAGCTC 1442  
 DB AsnAsnProArgTyrrIlyGlyGlyAspLeuThrLeuTyrrAlaIleAsnLeuHisAsnVal 460  
 QY 1443 ACCAAGTACTGGCGGTACCTTATCTTTTCTAAACAAGAGTGAATTAATACCTTCTA 1502  
 DB ThrIlyTyrrLeuAspIlyLeuProIlyrrProPheSerAsnIlyGlnIlyValAspIlyTyrrLeu 480  
 QY 1503 AGACTTTGGGACCTCATGATTAATTTTCAAACTGTGTCGAATGATGCTTAACCTTCTA 1562  
 DB ArgProLeuGlyProHisGlyLeuLeuSerIlySerValGlnLeuAsnGlyLeuThrIleu 500  
 QY 1563 AAGATGGTGAATGATCAAACTTGGCACCTTTAATGAAGAAAACCTCCGCGCGAGAGT 1622  
 DB LysMetValAspAspGlnThrLeuProProLeuMetGlnIlyProLeuArgProGlySer 520  
 QY 1623 TCATCGGCTGGCCGCTTTCATATAGTTTGTGATTAAGAAATGCAAGAGTCTCT 1682  
 DB SerLeuGlyLeuProAlaPheSerTyrrSerPhePheValIleArgAsnAlaIlyValAla 540  
 QY 1683 GCTTGCAATC 1691  
 DB 541 AlaCysIle 543  
 DB  
 RESULT 14  
 ABP56822  
 ID ABP56822 standard; protein; 545 AA.  
 AC ABP56822;  
 DT 02-APR-2003 (first entry)  
 XX  
 DB Human heparanase protein SEQ ID NO:18.  
 XX  
 KW Human; heparanase; phosphorothioate; antisense oligonucleotide;  
 KM Cystostatic; gene therapy; tumour; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003004705-A1.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 01-JUL-2002; 2002MO-US020636.  
 XX  
 PR 05-JUL-2001; 2001US-00899440.  
 XX  
 PA (UNCO ) UNIV COLUMBIA NEW YORK.  
 XX  
 PI Stein C;  
 XX  
 DR WPI; 2003-201558/19.  
 DR N-PSDB; AB222816.

XX New oligonucleotide having a sequence complementary to a sequence of  
PT ribonucleic acid encoding a heparanase, useful for preparing a  
PT composition for treating tumor.  
XX  
PS Disclosure; Page 46-47; 48pp; English.  
XX  
CC The present invention describes an oligonucleotide having a sequence  
CC complementary to a sequence of ribonucleic acid encoding a heparanase.  
CC The oligonucleotide hybridizes with the ribonucleic acid under conditions  
CC of high stringency and has a sequence comprising 10-40 bp. The  
CC inter-nucleotide linkages of the oligonucleotide comprise at least one  
CC phosphorothioate linkage. Hybridisation of the oligonucleotide to the  
CC ribonucleic acid inhibits expression of the heparanase, where inhibition  
CC of heparanase means at least a 50% reduction in the quality of  
CC heparanase in a cell; (2) a method of inhibiting expression of a  
CC oligonucleotide in an amount effective to inhibit the expression of a  
CC heparanase in the cell and a carrier; and (3) a method of treating a  
CC tumour in a subject comprising administering to the subject an amount of  
CC the above oligonucleotide effective to inhibit expression of a heparanase  
CC in the subject. Heparanase antisense oligonucleotides have cytostatic  
CC activity, can be used in gene therapy, and can be used for preparing a  
CC composition for treating tumours. The present sequence represents human  
CC heparanase, which is given in the exemplification of the present  
CC invention  
XX  
SQ Sequence 545 AA:  
  
Alignment Scores:  
Pred. No.: 3,676-270 Length: 545  
Score: 2817.00 Matches: 542  
Percent Similarity: 99.63% Conservative: 1  
Best Local Similarity: 99.45% Mismatches: 0  
Query Match: 90.32% Indels: 2  
DB: 6 Gaps: 2  
  
US-10-676-079-3 (1-1721) x ABP56822 (1-545)  
QY 63 ATGCTGCTGCGCTGAAGCCTGCGCTGCGCGCGCG---CTGATGCTGCTGCTGCGG 119  
Db 1 MetLeuLeuArgSerLysProAlaLeuProProProLeuLeuMetLeuLeuGly 20  
QY 120 CCGGTGGGTCCCTCTCCCTCGGCGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 176  
Db 21 ProLeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAspVal 40  
QY 177 GTGACCTGACCTTCTCACCGACGACCGCTGACCTGACCGCTGCTGCTGCTGCTGCTG 236  
Db 41 ValAspLeuAspPhePheThrGlnGluProLeuHisLeuValSerProSerPheLeuSer 60  
QY 237 GTCACTTGAACGCAACCTGGCGACGACCGCGGCTTCTGATCTCTGGGTTCTCA 296  
Db 61 ValThrIleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuGlySerPro 80  
QY 297 AAGCTTGTAACCTTGGCGACGAGCTTGTCTCTGCGTACCTGAGAGTTGGTGGACCAAG 356  
Db 81 LysLeuArgThrLeuAlaArgGlyLeuSerProAlaThrLeuAspPheGlyGlyThrLys 100  
QY 357 ACAGACTCTCTAATTTTCGATCCCAAGAGGAATCAACTTGAAGAGAGAGTTACTG 416  
Db 101 ThrAspPheLeuIlePheAspProLysGlySerThrPheGlnGluArgSerTyrTrp 120  
QY 417 CAATCTCAAGTCAACAGATATTTGCAATATGATCCATCCCTCTGATGTGGAGAG 476  
Db 121 GlnSerGlnValAsnGlnAspIleCysLysTyrGlySerIleProProAspValGlnGlu 140  
QY 477 AAGTTACGTTGGAATGCGCTTACACAGAGCAATTTGCTACTCCGAGAACCTACAGAA 536  
Db 141 LysLeuArgLeuGlnTyrProGlyGlnGlnGlnLeuLeuLeuArgGlnHisTyrGlnLys 160  
QY 537 AAGTTCAAGAACGACACTACTCAAGAGCTGTAGTGTGATACCTTTTGAAC 596

Db 161 LysPheLysAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsn 180  
QY 597 TCGTCAGACTGACCTTATGCTTGGCTTAATGCTTATTAAGAACAGACGAGTTGACG 656  
Db 181 CysSerGlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGln 200  
QY 657 TGGACAGTTCTAATGCTCAGTTGCTCTGAGACTACTGCTCTTCGAGGGGTAACTT 716  
Db 201 TrpAsnSerSerAsnAlaGlnLeuLeuLeuAspTyrCysSerSerLysGlyTyrAsnIle 220  
QY 717 TCTTGGAACTAGACATGAACCTAACAGTTTCTTAAGAGGCTGATATTTTCATCAT 776  
Db 221 SerTrpGluLeuGlnGlnGluProAsnSerPheLeuLysValAlaAspIlePheIleAsn 240  
QY 777 GGGTCGAGTTAGAGAGATTATATTCATATTCATTAACCTTGAAGAAAGTCACCTTC 836  
Db 241 GlySerGlnLeuGlyGluAspPheIleGlnLeuHisLysLeuLeuArgLysSerThrPhe 260  
QY 837 AAAAAATGCAAACTCTATGCTCTGATGTTGGTTCAGCTTGAAGAAAGCGCTTAAGATG 896  
Db 261 LysAsnAlaLysLeuTyrGlyProAspValGlyGlnProArgArgThrAlaLysMet 280  
QY 897 CTGAAGAGCTTCTGAAGGCTGGTGGAGAAAGTATGATTCAGTTACATGACATCATAC 956  
Db 281 LeuLysSerPheLeuLysAlaGlyGlyGluValIleAspSerValThrTrpHisIleTyr 300  
QY 957 TATTTGAATGACGAGCTGCTACACAGAGAAAGATTTTCTAAACCTGATGATTGACATT 1016  
Db 301 TyrLeuAsnGlyArgThrAlaThrArgGluAspPheLeuAsnProAspValIleAspIle 320  
QY 1017 TTTATTTTCATCTGTGCAAAAAGTTTCCAGGTGTTGAGACGCCAGGCTGCGCAAGAG 1076  
Db 321 PheIleSerSerValGlnLysValPheGlnValValGlnSerThrArgProGlyLysLys 340  
QY 1077 GTCTGGTTAGAGAAACAAGCTGATATGAGAGCGAGCGCGCTGCTATCCGACACC 1136  
Db 341 ValTrpLeuGlyLysThrSerSerAlaTyrGlyGlyAlaLeuProLeuLeuSerAspThr 360  
QY 1137 TTTGACGCTGCTTATATGCTGATTAATTTGGGCTGTACGCGCGAATGGGAATAGA 1196  
Db 361 PheAlaAlaGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaArgMetGlyIleGlu 380  
QY 1197 GTGATGATGAGGAATATTTCTTTGAGACGAGAACTACATTAATGATGAGAACTTC 1256  
Db 381 ValValMetCysGlnValPhePheGlyAlaGlyAsnTyrHisLeuValAspGluAsnPro 400  
QY 1257 GATCTTTTACCTGATTAATTTGCTATCTCTTCTTGAAGAAATTTGGTGGGACCAAGTG 1316  
Db 401 AspProLeuProAspTyrTrpLeuSerLeuLeuPheLysLysLeuValGlyThrLysVal 420  
QY 1317 TTAATGGCAAGCGTGCAGAGTTCAAAAGAGAGAGCTTCAGATATACCTTCATTCACA 1376  
Db 421 LeuMetAlaSerValGlnGlySerLysArgArgLysLeuAspValTyrLeuHisCysThr 440  
QY 1377 AACACTGACAATCCAGATTAATAAGAGAGATTAACCTGTGATGCAATTAACCTCCAT 1436  
Db 441 AsnThrAspAsnProArgTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLysHis 460  
QY 1437 AACGTCAACAAGTACTGGGTTACCTTATCCCTTTTCTTAACAAGCAAGTGAATTAATAC 1496  
Db 461 AsnValThrLysTyrLeuArgLeuProTyrProPheSerAsnLysGlnValAspLysTyr 480  
QY 1497 CTTCTAAGACTTTGGGACCTCATGATTAATTTTCAAAATCTGTCCAACCTCAATGCTTA 1556  
Db 481 LeuLeuArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeu 500  
QY 1557 ACTCTTAAGATGGTGAATGATCAAACTTGGCACCTTTAATGAGAAAACTCTCCGGCCA 1616  
Db 501 ThrLeuLysMetValAspAspGlnThrLeuProProLeuMetLysLysProLeuAspPro 520  
QY 1617 GGAAGTTCACTGGGCTTGGCAGCTTCTCATATAGTTTGTGATTAAGAAATGCGCAA 1676  
Db 521 GlySerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLys 540





Db 161 LysPheLeuAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsn 180  
QY 597 TGCTCAGAGACTGGACTTATCTTTGGCCTTAATCGTTATTTAGAGACGAGATTTCAG 656  
Db 181 CysSerGlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGln 200  
QY 657 TGGACAGCTTCTATGCTCAGTTGCTCTGAGCTACTGCTCTTCCAAAGGGGTATTAACATT 716  
Db 201 TrpAsnSerSerAsnAlaGlnLeuLeuAsnProTyrCysSerSerLysGlyTyrAsnIle 220  
QY 717 TCTTGGGAAGCTAGGCAATGAACCTTAACAGTTTCCTTAAAGAGGCTATTTTCATCAAT 776  
Db 221 SerTrpIleuGlnLysGlnProAsnSerPheLeuLysValAspIlePheIleAsn 240  
QY 777 GGGTCGAGTTAGAGAGATTAATTCATTCATTAACCTTCTAAGAAAGCTCAACCTTC 836  
Db 241 GlySerGlnLeuGlyLysAspPheIleGlnLeuHisLysLeuLeuArgLysSerThrPhe 260  
QY 837 AAAAAATGCAAAATCTATGCTCTGATGTTGGTCAAGCTCGAAGAAAGCGCTTAAGATG 896  
Db 261 LysAsnAlaLysLeuTyrGlyProAspValGlyGlnProArgGlyThrAlaLysMet 280  
QY 897 CTGAAGAGCTTCTGAAGGCTGTGAGAGATGATGATTGATTCACTACATGCGCATCACTAC 956  
Db 281 LeuLysSerPheLeuLysAlaGlyGlyGlnValIleAspSerValThrTrpHisIleTyr 300  
QY 957 TAATTGAATGACGAGACTGTCTACAGGAGAAATTTCTAAACCTGATGTAATGGACATT 1016  
Db 301 TyrLeuAsnGlnArgThrAlaThrArgLysPheLeuAsnProAspValLeuAspIle 320  
QY 1017 TTTATTTCACTGTGCAAAAAAGTTTTCAGAGTGTGAGACACAGCCCTGCGCAAGAG 1076  
Db 321 PheIleSerSerValGlnLysValPheGlnValAlaGlnSerThrArgProGlyLys 340  
QY 1076 GTCTGTTAGAGAAACAAGCTCTGATATGAGAGGCGAGCGCCCTTGCTATCCGACACC 1136  
Db 341 ValTrpLeuGlyGlnThrSerSerAlaTyrGlyAlaProLeuLeuSerAspThr 360  
QY 1137 TTTCGAGCTGGCTTTATGTGGCTGGATTAATTTGGGCTGTGACCCGCAATGGGAATGAA 1196  
Db 361 PheAlaIleGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaArgMetGlyIleGln 380  
QY 1197 GTGGTGAAGGCAAGATTTCTTTGAGCAGAGAACTAACCTTTAGTGATGAAAACTTC 1256  
Db 381 ValValMetArgGlnValPhePheGlyAlaGlyAsnTyrHisLeuValAspGluAsnPhe 400  
QY 1257 GATCCTTACCTGATTAATTTGGCTATCTCTTCTGTTCAAGAAATTTGGGGCAACCAAGTG 1316  
Db 401 AspProLeuProAspTyrTrpLeuSerLeuLeuPheLysLysLeuValGlyThrLysVal 420  
QY 1317 TTAATGGCAAGCTGCAAGGTTCAAGAGAAAGAAAGCTTCGAGTAACTTCATTGCA 1376  
Db 421 LeuMetAlaSerValGlnGlySerLysArgLysLeuArgValTyrLeuHisCysThr 440  
QY 1377 AACACGTACAATCCAAAGTATTAAGAAGAGATTAATCTGTAATGSCATTAACCTCCAT 1436  
Db 441 AsnHisArgAsnProArgTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLeuHis 460  
QY 1437 AACGTCAACAAGTACTTGGGTTACCTTACCTTCTTTCTAACAAGCAAGTGATTAATAC 1496  
Db 461 AsnValThrTyrTyrLeuAspGlnLeuProTyrProPheSerAsnLysGlnValAspLysTyr 480  
QY 1497 GTTCTAAGACCTTTGGAGCTCATGATTAATTTCCAAATCTGTCAACTCATATGTTCTA 1556  
Db 481 LeuLeuArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeu 500  
QY 1557 ACTCTAAGATGGTGAATGAATCAAACTTGCCACCTTAATGAAAAAAGCTCCGGGCA 1616  
Db 501 ThrLeuLysMetValAspAspGlnThrLeuProLeuMetGlnLysProLeuAspPro 520  
QY 1617 GGAAGTTCACTGGGCTTGCAAGCTTTCTCATATAGTTTTTTTGGATAGAAGAAATGCCAA 1676  
Db 521 GlySerSerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLys 540

QY 1677 GTTGTGCTTGATC 1691  
Db 541 ValAlaIleCysIle 545

Search completed: August 28, 2004, 07:58:11  
Job time : 233 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 28, 2004, 07:42:36 ; Search time 65.5 Seconds

(without alignments)  
5054.828 Million cell updates/sec

Title: US-10-676-079-3

Perfect score: 3119

Sequence: 1 ctgaagcttcctgactctcg.....atactagctgctgactg 1721

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xlp  
-O=/cgn2.1/USFTO\_spool\_p/US10676079/runat\_28082004\_084226\_23829/app\_query.fasta\_1.1863  
-DB=PIR\_78 -QFMT=fastan -SUFIX=tpir -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10676079 @CGN 1.1.89 @runat\_28082004\_084226\_23829 -NCP=6 -ICU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFBLLOC=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: PIR\_78: \*  
2: pir1: \*  
3: pir2: \*  
4: pir3: \*  
5: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897.5	28.8	480	2	UC7506
2	416	13.3	521	2	UC7506
3	169.5	5.4	190	2	UC7506
4	135	4.3	411	2	T01953
5	124.5	4.0	1044	2	S74760
6	118.5	3.8	2298	2	T10050
7	113.5	3.6	287	2	T49648
8	112.5	3.7	314	2	T65765
9	112.5	3.6	356	2	UC5273
10	112	3.6	688	2	F64383
11	112	3.6	1048	2	S32961
12	111.5	3.6	575	2	A27421
13	111	3.6	463	2	T12094
14	111	3.6	670	2	T13425
					T10666

15	110.5	3.5	503	1	LURB11	annexin XI - rabbi
16	110	3.5	607	2	AH2891	thiamin biosynthes
17	110	3.5	644	2	D97667	thiamin biosynthes
18	109.5	3.5	505	2	A53152	annexin XI - human
19	109.5	3.5	879	2	E91031	probable outer mem
20	109	3.6	627	2	A44112	spidroin 2, dragli
21	108	3.6	460	2	T33110	hypothetical prote
22	108	3.6	1500	1	JQ1348	carbamoyl-phosphat
23	107.5	3.4	500	2	D87541	beta-xylosidase li
24	107.5	3.4	738	2	T00343	hypothetical prote
25	107	3.4	1247	2	T42209	neural plakophilin
26	106	3.5	403	2	A53662	homeotic protein H
27	106	3.4	879	2	F85875	probable fibritin
28	106	3.4	1579	2	S59801	protein kinase SSK
29	105.5	3.4	511	2	S61166	probable membrane
30	105	3.4	670	2	T38446	microtubule-associ
31	105	3.4	1048	2	T31425	C-terminal domain
32	105	3.4	2013	2	A11489	probable peptidogl
33	104.5	3.4	788	1	S00652	phosphoribosylam
34	104.5	3.4	847	2	AG1001	nitrite reductase
35	104.5	3.5	1215	2	T32734	myosin-1A - Acant
36	104	3.3	402	2	H97114	uncharacterized co
37	104	3.3	432	2	F70411	adenylosuccinate s
38	104	3.3	1572	2	S45251	SNF2alpha protein
39	103.5	3.4	228	2	A44982	collagen UCO1 - p
40	103.5	3.3	587	2	S36231	beta-fructofuran
41	103.5	3.3	676	2	AF1153	transcription anti
42	103.5	3.3	687	2	F65188	retrotransposon li
43	103	3.3	796	2	D97065	transketolase (imp
44	103	3.3	892	2	T08973	probable serine/ch
45	103	3.3	1266	2	AF0911	probable exported

#### ALIGNMENTS

RESULT 1  
UC7506  
heparanase protein 2a - human  
C/Species: Homo sapiens (man)  
C/Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 01-Dec-2000  
C/Accession: UC7506  
R/McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.; Pat  
Blochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A/Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me  
A/Reference number: UC7506  
A/Accession: UC7506  
A/Molecule type: mRNA  
A/Residues: 1-480 <MCK>  
A/Cross-references: GB:AF282885  
C/Comment: This protein, a intracellular membrane-bound enzyme, has biological and therai  
therapies.  
C/Genetics:  
A/Gene: hpa2a  
A/Map position: 10q23-10q24  
C/Keywords: heparin binding; membrane bound

Alignment Scores:	Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-10-676-079-3 (1-1721) x UC7506 (1-480)		5,076-60	480	202	74	146	139	9
QY	120	CCGCTGGGTCCTCCCTGCGCCCTG	149					
DB	16	ProPROMA1CySleuAlaProGlyAlaLeuYrleuAlaLeuLeuHleuSerleu	37					
QY	150	-----CCCGACCTGCG	167					
DB	38	SerSerGlnAlaGlyAspArgArgProLeuProValaAspArgAlaGlyLeuYsglu	57					

```

QY 168 CAGACAGCTGCGAGCTGGAATTCTTCAACCCGAGCGCGCTGACCTGTGAGCCCTCG 227
   ::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 LyrThrLeuLeuLeuAerValSerThrLysAsnProValArgThrValAsnGluAsn 77
QY 228 TTCTGTCTCCACCATTTGACGCCACCTGGCCAGCGCGGTTCTTCATCTCTCTG 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 PheLeuSerLeuGlnLeuAerProSerLeuLeuHisAsp--GlyTrpLeuAerPheLeu 96
QY 288 GATTCTCCAAAGCTTCGTACCTTGCCAGAGGCTTGCTCTGCGCTGACCTGAGGTTGGT 347
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 SerSerLysArgLeuValThrLeuAlaArgGlyLeuSerProAlaPheLeuArgPheGly 116
QY 348 GGCACCAACAGACACTTCTTAATTTTC-----GATCCCAAGAAAGATCAACC 395
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 GlyLysArgThrAerPheLeuGlnPheGlnAsnLeuArgAsnProAlaLysSerArg--- 135
QY 396 TTGGAAGAGAAAGTACTGGCAATCTCAAGTCAACAGATATTGCAAAATATGATCC 455
   136 -----GlyGly 137
QY 456 ATCCCTCTGATGTGAGAGAAAGTTACGGTTGGATGCCCTTACAGAGCAATTGCTA 515
   |||||
Db 138 ProGlyProAsp----- 141
QY 516 CTCGAGAACACTTACAGAAAAAGTTCAAGAACAGACCTTACAGAGAGCTGTAGAT 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 142 -----LyrTrpLeuLysAsnTrpGlu----- 148
QY 576 GTGCTATACACTTTTGCAAACTGCTCAGAGACTGACCTTGATCTTTGGCTTAATGCTTA 635
   148 ----- 148
Db 636 TTAAGAACAGACAGATTGTCAGTGAACAGTTCTAATGCTCAGTTGCTCTGACTACTGC 695
   148 ----- 148
QY 696 TCTTCCAAAGGGTAAACATTCTTGGGAACAGCAATGAACCTTACAGTTCCCTTAAG 755
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 -----AspGlnProAsnAsnTrpArgThr 156
QY 756 AAGCGTATATTTTCATCAATGGTGGCAGTTAGAGAAATTTATATTCATTCATTA 815
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 MetHisGlyArgAlaValAsnGlySerGlnLeuGlyAspTrpTrpGlnLeuLysSer 176
QY 816 CTTCCTAAGAAAG--TCCACCTTCAAAAATGCAAAACTATAGTGTCTGATGTGTCAG 872
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 LeuLeuGlnProLeuArgGlyLeuSerArgAlaSerLeuTrpGlyProAsnIleGlyArg 196
QY 873 CTTGGAAGAAAGACGCTTAAGATGCTGAAGACTTCTGGAAGCTGTGGAAGATGATT 932
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 197 ProArgLysAsnValIleAlaLeuLeuAspGlyPheMetLysValAlaGlySerThrVal 216
QY 933 GATTCACTTACAGCGCTCACTACTATTTGAATGAGAGAGCTGTACCAAGAAATTT 992
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 AspAlaValThrTrpGlnHisCysTrpLysAspLysArgValValLysValMetAspPhe 236
QY 993 CTAAACCCCTGATGATGACATTTTATTTTCACTCTGCAAAAAGTTTCCAGTGCTT 1052
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 LeuLysThrArgLeuLeuAerPthrLeuSerAspGlnIleArgLysIleGlnLysValVal 256
QY 1053 GAGAGACACAGCGCTGGCAAGAAAGTGTGTTAGAGAAACAAGCTGTGATATGAGGC 1112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 257 AsnThrLysThrProGlyLysLysIleTrpLeuGlnLysValValThrTrpSerAlaGly 276
QY 1113 GAGAGCCGCTTGTATCCGACACTTTGACGCTGGCTTTATGTGCGCTGATAAATGGGC 1172
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 GlyThrAsnAsnLeuSerAspSerTrpAlaAlaGlyPheLeuTrpLeuAsnThrLeuGly 296
QY 1173 CTGTCAAGCCCAATGGAATAGAAAGTGTGATGAGCAAGATTTCTTTGAGCAGAGAAC 1232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 MetLeuAlaAsnGlnGlyIleAspValValIleArgHisSerPheAspPheArgLysTyr 316

```

```

QY 1233 TACCAATTAGTGATGAAAACTTGATCTTTTACCTGATTAATGATGCTATCTTCTGTC 1292
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 317 AsnHisLeuValAspGlnAsnPheAsnProLeuProAspTrpTrpLeuSerLeuLeuLys 336
QY 1293 AAGAAATTTGGTGGGACCAAGGTTTATATGCGAAGGTCGCAAGGTTCAAGAAAG--- 1349
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 337 LysArgLeuIleGlyProLysValLeuAlaValHisValAlaGlyLeuGlnArgLysPro 356
QY 1350 -----AAGCTTCGAGTATACCTTTCATTCATTCGACAAACACTGAC 1385
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 ArgProGlyArgValIleArgAspLysLeuValGlyIleTrpAlaHisCysThrAsnHisLys 376
QY 1386 AATCCAAAGGATTAAGAGAGAGATTTTAACTGTGTATGCAATAAACCTCCATTAAGTCAACC 1445
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 AsnHisAsnTrpValArgGlySerIleThrLeuPheIleIleAsnLeuHisArgSerArg 396
QY 1446 AACTACTGCGGTTACCCCTATCTTTTTCATACCAACAGTGATTAATACCTTTCATAA 1505
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 397 LysLysIleLysLeuValaGlyThrLeuArgAspLysLeuValHisGlnTrpLeuLeuGln 416
QY 1506 CCTTTGGGACCTTCATGATTAATTTCCAAATCTGTCCAATGCTCAATGCTTAATG 1565
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 ProTrpGlyGlnGlnLysLysSerLysSerValGlnLeuAsnGlyGlnProLeuVal 436
QY 1566 ATGTGAGATGATCAACCTTGGCCACTTAAATGAAAAAAGCTCTCGGCCAGAGACTCA 1625
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 437 MetValAspAspGlyThrLeuProGlyLeuLysProArgProLeuArgAlaArgThr 456
QY 1626 CTGCGCTTCGCGACTTTCATATACTTTTGTGATTAAGAAATCCCAAGTGTGCT 1685
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 LeuValIleProProValThrMetGlyPhePheValValLysAsnValAsnAlaLeuAla 476
QY 1686 TGC 1688
   |||
Db 477 Cys 477

RESULT 2
T45608
hypothetical protein F13G24.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45608
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.-J.; Voet, M.; Robben, J.; Volckaert, G.; Bar
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23009
A:Accession: T45608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <BEV>
A:Cross-references: EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3
A>Note: F13G24.30

Alignment Scores:
Pred. No.: 1,78e-23 Length: 521
Score: 416.00 Matches: 154
Percent Similarity: 42.05% Conservative: 68
Best Local Similarity: 29.17% Mismatches: 184
Query Match: 13,344 Indels: 122
DB: 2 Gaps: 24

US-10-676-079-3 (1-1721) x T45608 (1-521)
QY 285 CTGGGTTCTCCAAAGCTTGTGACCTTGGCCAGAGCTTGTCTCTGCTGACCTGATT 344
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 LeuThrArgProLeuLeuThrLysAlaIleLysAlaPheLysProLeuArgLysIle 74
QY 345 GGTGGACCAACAGACACTCTCTAATTTTCATTCACCAAGAAAGATCAACCTTTGAAGAG 404
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 GlyGlySerLeuGlnAspGlnValIleTrpAspValGlyAsnLeuLysThr----- 91

```

```

QY 405 AGAAGTACTGGCAATCTCAAGTCAACGAGATATTGGCAATATGATCCATCCCTCT 464
DB 92 -----ProGlyArg----- 94
QY 445 GATGTGGAGAGAGATTACGGTTGGAAATGGCCCTTACCGAGAGCAATGGTACTCCGAGAA 524
DB 95 -----PropheGlnLysMet----- 99
QY 525 CACTACACGAAAAGTTCAAGAACAGC-----ACTTACTCAAGAAAGCTCTGTGA--- 572
DB 100 -----AsnSerGlyLeuPheGlyPheSerLysGlyCysLeuHis 112
QY 573 -----GATGTGCTATATACATCTTGGCAAACTGCTCAGAGCTGGACTGGATGATCTT 620
DB 113 MetLysArgTTPAspGlnLeuLysAsnSerPheLeuThrAlaThrGlyAlaValValThrPhe 132
QY 621 GGCCTAAATGCGTTATTTAAGAACAGACAGATTGACG-----TGG 659
DB 133 GlyLeuAsnAlaLeuArgGlyLysArgHisLysLeuArgGlyLysAlaAlaTrpGlyLysAlaTrp 152
QY 660 AACAGTTCTTAATGCTCAGTTCCTCGAGCTACTGCTCTTCCAAAGGGGTATTAACAT--- 716
DB 153 AspHisLysLeuThrGlnAspPheLeuAsnTyThrValSerLysGlyTyThrValIleAsp 172
QY 717 TCTTGGGAATGAGCAATGAACTTAACAGTTTCTTAAGAAAGGCTGATATTTTCATCAAT 776
DB 173 SerTrpGlnPheGlyAsnGlnLeuSerGly-----SerGlyValGlyLysAlaSerValSer 190
QY 777 GGGTGGCAGTTAAGAGAGATTATATTCATATTCATTAACCTTCTAAGAAAGTCCACTTC 836
DB 191 AlaGlnLeuTyrglyLysAspLeuIleValLeuLysAspValIleAsnLys---ValTy 209
QY 837 AAAAAATGCAAAATCTCATATGCTCTCATGTTGCTCAGCCT-----CGAGA 881
DB 210 LysAsnSerTrpLeuHisLysAspProIleLeuValAlaProGlyGlyPheTyrglnGlnGln 229
QY 882 AAGACCGCTAAGATGCTGAAGAGCTTCTGAAGGCTGTGGAGAGATGATGATTCAGT 941
DB 230 TrpTyThrLysLeuLeuGlnLys-----SerGlyProSerValValAspValVal 246
QY 942 ACATGGCATCTACTATATTGAATGAGCAGACT-----GCTACCGAGAGATTTTCTA 995
DB 247 ThrHisLysIleTyraAsnLeuGlySerGlyAsnAspProAlaLeuValLysLysIleMet 266
QY 996 AACCTCGATGTATGAGCACTTTTATTTATTCCTGTGCAAAAAGTTTC----- 1043
DB 267 AspProSer-----TyrLeuSerGlnValSerLysThrPheLysAspValAsn 282
QY 1044 CAGGTGGTTGAGAGCACAGGCTGGCAGAAAGGTCTGGTTAGGAGAAACAAGCTCTGCA 1103
DB 283 GlnThrIleGlnGlnHisGlyProTrpAlaSerProTrpValGlyLysSerGlyAla 302
QY 1104 TATGAGGCGGAGCGCCCTGCTATCCGACACTTTCAGCTGGCTTATGTCGCTGAT 1163
DB 303 TyrAsnSerGlyGlyArgHisValSerAspThrPheIleAspSerPheTrpTyrLeuAsp 322
QY 1164 AATTGGGCTGTGACCCCGAATGGGAATGAAGTGTATGAGGGAAGTATTTCTTGGCA 1223
DB 323 GlnLeuGlyMetSerLysAlaArgHisAsnThrLysValTyCysArgGlnThrLeuValGly 342
QY 1224 GGAGGAAATACCATTTAGTGAGATGAA---AAGTTGATCGCTTATCTGATTTAGGCTA 1280
DB 343 ---GlyPheTyrglyLeuLeuGlnLysGlyThrPheValProAsnProAspTyTrpSer 361
QY 1281 TCTCTTCTGTCAAGAAATGTGTGGCAGCAAGGTGTATATGGCAGCGTGCAGGTTCA 1340
DB 362 AlaLeuLeuTrpHisArgLeuMetGlyLysGlyValLeuAlaValGlnThrAspGlyPro 381
QY 1341 AAGAGAAAGAGTTGAGATATCTTCATGTGCAAAAACATGACATCAAGTATTA 1400
DB 382 Pro-----GlnLeuArgValTyraLeuHisCysSerLys-----Gly 393

```

```

QY 1401 GAAGGAGATTTAACTGTGTATGATCCATTAACCTCCATTAACGTCCACCAAGTACTG----- 1454
DB 394 ArgAlaGlyValThrLeuLeuLeuIleHisLeuSerAsnGlnSerAspPheThrValSer 413
QY 1454 ----- 1454
DB 414 ValSerAsnGlyIleAsnValValLeuAsnAlaGlnSerArgLysLysSerLeuLeu 433
QY 1455 ---CGGTTACCTTAATCTTTTCT-----AACAGCAAGAGGATTAATACCTTCTA 1502
DB 434 AspThrLeuLysArgProPheSerTrpIleGlySerLysAlaSerAspGlyTyrLeuAsn 453
QY 1503 AGACCT-----TTGGACCTCATGCA-----TTACTTTCAAATCTGTCCACTC 1547
DB 454 ArgGlnGlnTyrglyLeuLeuThrProGlnAsnGlyValLeuArgSerLysThrMetValLeu 473
QY 1548 AATGCTTAATCTTAAAGATGTGTGATATCAAACTTGGCCACTTATATGAAAAAAGT 1607
DB 474 AsnGlyLysSerLeuLysProThrAlaThrGlyAspIleProSerLeu---GluProVal 492
QY 1608 CTCCGGGCA---CGAAGTTCACCTGGCGCTTCTCATATATGTTTGTGATA 1664
DB 493 LeuArgSerValAsnSerProLeuAsnValLeuProLeuSerMetSerPheIleValLeu 512
QY 1665 AGAAATGCCAAAGTTGCTGCTTGC 1688
DB 513 ProAsnPheAspAlaSerAlaCys 520

RESULT 3
T01953
hypothetical protein T2L5.6 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 21-Jan-2000
C:Accession: T01953
R:Geisels, C.; Smith, A.; Le, T.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of A. thaliana T2L5.
A:Reference number: 214470
A:Accession: T01953
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-190 <GB>
A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:introns: 36/2; 69/3
A:Note: T2L5.6
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Alignment Scores:
Pred. No.: 6,86e-05 Length: 190
Score: 169.50 Matches: 54
Percent Similarity: 45.36% Conservative: 34
Best Local Similarity: 27.84% Mismatches: 57
Query Match: 5.43% Indels: 49
DB: 2 Gaps: 9

US-10-676-079-3 (1-1721) x T01953 (1-190)
QY 1206 AGCAAGATTTCTTGGAGACAGAACTACCATTTAGTGAT---GAAGACTTCGATCCT 1262
DB 12 ArgGlnSerLeuIleGly---GlyAsnTyrglyLeuLeuAsnThrThrAsnPheThrPro 30
QY 1263 TTACTGATTTATGGCTATCTCTTCTGTTCAAGAAATGTGTGGCAGCAAGGTGTATATG 1322
DB 31 AsnProAspTyTrpSerAlaLeuIleTrpArgGlnLeuMetGlyArgLysAlaLeuPhe 50
QY 1322 GCAAGCGTGCAGGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382
DB 51 ThrThrPheSerGlyThrLys-----LysIleLeuSerTyThrHisCysAla----- 66
QY 1383 GACAATCCAGAGTATTAAGAGAGAGATTTAACTCTGTATGCAATTAACCTCCATTAACGTC 1442

```

[illegible]

RESULT 4  
S74760  
hypothetical protein slr1617 - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S74760  
O:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 103-116, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
s.  
A:Reference number: S74332; MUID:97061201; PMID:8905231  
A:Accession: S74760  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-411 <KAN>  
A:Cross-references: EMBL:D90901, GB:AB001339, NID:G1651897, PIDN:BAAL6911.1, PID:d101764  
/Note: the nucleotide sequence was submitted to the EMBL Data library, June 1996

Alignment Scores:	
Pred. No.:	0.035
Score:	135.00
Percent Similarity:	39.46%
Best Local Similarity:	21.84%
Query Matchn:	4.33%
DB:	2
Length:	411
Matches:	89
Conservative:	72
Mismatches:	122
Indels:	125
Gaps:	22

[illegible]

QY	479	GTTCAGCTTGGAATGGCCCTTACCGAGCAATCTCTACTCCGAGAACACTACGAAAAA	538
Db	137	GIyPhetiegiYhrrhlaLeuGlnGlyAla-Leuile-----AamSerg1	151
QY	539	GTTCAGAACAAACGACCTACTCAAAAGAGCTGTGATGTCTATACACTTTTGGCAACTG	598
Db	151	uPhetArgLeuileSerProthrrArgGlnGlnIleAsple-----PhehlaGlyse	168
QY	599	CTCAGAGACTGACCTTGATCTTT-----GGCCTAAATCCGTTATTAAAGACACGA	649
Db	168	rThrlYsLeuAspLeuAlaSerGlnGluAsnIleAspCysrileValIleLeuAlaAs	188
QY	650	TTTGAGAGGAAACGTTCTAAATGCTCAGTTG-----CT	682
Db	188	nProArgValYrthrrSerAsnValAlaIleMetGlyGlnThrLeuthrMetLeuArgAsnVa	208
QY	683	CCTGGACTACTGCTCTTCCAAAGGG-----TATAACATTTCTTGGACATA--	728
Db	208	IlleAspValCysLeuAlaYsAspleProLeuIleYrProSerSerTrpGlnIleYr	228
QY	729	-----GGCAATGAACCTAACAGTTTCCTTAAGAG--	758
Db	228	rSergIlyTyralaGlyThrIleAlaAspGlnSerThrProAlaLeuProArgGlyPr	248
QY	759	-----GCTGATATTTTCAATCAATGAGTGCAGTTAGAGAGA	796
Db	248	oTyrcIyGlnThrYsTrpLeuAlaGlnIleLeuile-----As	261
QY	797	TTATATTCAATTGCATATACTTCTTAAGAAATCCACTTCAAAATGCAAAATCTGTATG	856
Db	261	PhicysrAspArgThrArgGlyLeuArgCysAlaIleLeuArgSerSerProValYrG1	281
QY	857	TCCGATGTTGTGAGCCCTGCAGAAAGACGGCTAACAGTGCAGAAAGCCTTCCGAAGGC	916
Db	281	ySerMetSerAspYsPro-----LysPheIlePheAsnPheIleYsly	296
QY	917	TGTGTGAGAACTGATGTTGATTCAGTTAATGAGCATCACTACTATTGTGAATGAGCAGATGC	976
Db	296	sAlaSerGlnIyGlnIlySileValThr-----HisIleYrIleAsngly-----	311
QY	977	TACCAAGGAGATTTTCTAAACCTGATGTA-----TTGACATTTTATTTTC	1024
Db	312	-----AenProIlySleAsnAspLeuAsnIleAspAspleuIleSe	325
QY	1025	ATCTGTGCAAAAAGTTTCCAGGTGGTGTGAGAGACACGAGCCTGCAGAAAGATCTGGTT	1084
Db	325	rSerIleValAlaThrLeu-----LysSerArgPheI1	336
QY	1085	AGGAGAAACAGCTCTGCATATYGAGGCGAGCGCCCTTGCTATCCGACACTTTCAGC	1144
Db	336	egIyAsnLeuAsnIle-----GlyThrGlyGlnLeuSerSerThrLeuYrI1	352
QY	1145	TGCGCTTATGCGCTGATATAATTGGCCCTGACGC-----CGAATGG	1189
Db	352	eAlaGlnMetIleArgAspGlnLeuGlySerSerSerMetIleGlnGlnIleGluValAs	372
QY	1190	AATAGAAGTGTGTAGAGCAAGATTTCTTTGGAGACAGAAACTACATTAGTGAATGA	1249
Db	372	nThcGlnValAlaSerIleAlaMetAsnYrGlyArgAlaAsn--HisValIleuAsp--	390
QY	1250	AAACTGCAATCTTTAACCTGATTATGGCTATCTCTGTGTTCAAGAAATTTGGTGGCAC	1309
Db	391	-----TrpGlnProValIlePhePheGlnu--GlnGlyLe	401
QY	1310	CAAGGTGTTATGGCAAGCGTG	1331
Db	401	uYsSerLeuLeuIleGlnIle	408

RESULT 5  
T10050  
integrin alpha-v chain precursor - mouse  
N:Alternate names: vitronectin receptor alpha chain  
C:Species: Mus musculus (house mouse)

C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
 C/Accession: T10050  
 R/Wade, J.; Kumar, A.; Liu, Z.; Ruoslahti, E.; Reichardt, L.; Marvaldi, J.; Kanwar, Y.S.  
 J. Cell Biol. 132, 1161-1176, 1996  
 A>Title: Cloning of mouse integrin alpha v cDNA and role of the alpha v-related matrix r  
 A/Reference number: Z16920; M01D:96176109; PMID:8601592  
 A/Accession: T10050  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1044 <MAD>  
 A/Cross-references: EMBL:U14135; NID:9537490; PIDN:AAC52497.1; PID:9537491  
 A/Experimental source: strain CD-1, kidney  
 C/Superfamily: integrin alpha-2b chain  
 C/Keywords: cell adhesion; receptor

## Alignment Scores:

Score:	0.281	Length:	1044
Percent Similarity:	124.50	Matches:	137
Best Local Similarity:	35.978	Conservative:	86
Query Match:	22.108	Mismatches:	213
DB:	3.998	Indels:	184
	2	Gaps:	35

US-10-676-079-3 (1-1721) x T10050 (1-1044)

```

QY 48 CCAGGTGACCCCAAGTGTGCTGCGCTGCAAGCCTGCGCGCGCGCGCTGATGCTG 107
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4 ProGly-----ArgLeuLeuLeuLeuArgProArgProGly-----GlyLeuLeuLeu 18

QY 108 CTGCTCTCGGGCGCGCTGGGTGCTCTCTCCCTCGCGCGCGCTGCGCGCTGCGAAGA 167
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 19 LeuLeuProGlyLeuLeuLeuProLeuAla----- 28

QY 168 CAGGACGTGTGAGCCTGTGACTTTCTTACCACGAGCCGCTGCACCTGTGTAGCCCC 224
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 29 --AspAlaPheAsnLeuAsp-----ValGluSerProAlaGluTyrAlaGlyProGlu 45

QY 225 ---TCGTCCTGCTCCGCTGACATTTGAC---GCCAACCTGACCGACCGCGCGCTCTC 278
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 46 GlySerTyrPheGlyPheAlaValaAspPheGluProSerThrSerSerArgMetPhe 65

QY 279 ATCTCTCGGGTGTTCGAAGCTTGTACCTTGGCCAGAGGCTTGTCTCTGCTGACTG 338
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 66 LeuLeuValGlyAlaProGlyAlaAsnThrThrGlnProGlyIleValGlu----- 82

QY 339 AGGTGTGTGCGCACCAAGACACTTCTTAATTTTGATCCCAAGAAAGAAATCAACTTT 398
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 83 -----GlyGlyGlnValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 100

QY 399 GAAGAGAAAGTTACTGCAATCTCAAGTCAACCAAGATATTTGCAATATGATGATCATC 458
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 100 eGluPheAspSerThrGlyAsn-----ArgAspTyrAlaAlaAspAspPro-- 115

QY 459 CCTCTGATGTGAGAGAGAGTTACGGTGTGATGAGCCCTTACCAAGAGCAATTTGCTACTC 518
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 116 -----LeuGluPheLeuSerThrGlnThrPheGlyAla----- 126

QY 519 CAGAACTACTACCAAGAAAGTTCAAGAAAGCAAGCACTTCAAGAAAGCTCTGTAGTGTG 578
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 127 -----SerValArgSerLeuLeuGlnAsp-----LysValLeuAlaCysAs 139

QY 579 CT-----ATACACTTTTGCAAACTGCTGAGAGCTGAGATGATCTTTGCGCTAAATGCG 632
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 139 LaProLeuTyrThrIlePheGlnThrGluMetLeuGlnIleGluProValGlyThrCysP 159

QY 633 TTATTAAAGACAGCAGATTGTCAGTGAACAGTTCTTAATGC-----T 674
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 159 heLeuGlnIlePheGlyThrIleValGlyTyrAlaProCysArgSerLeuAsnIleAspA 179

QY 675 CAGTGTCTCTGCACTGCTCTTCCAGAGGCTTAATCAATTTCTTGGAACTAGCAAT 734
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 179 laAspIleGlnGlyPheCysGlnGly---GlyPheSerIleAspPhe----- 193
  
```

```

QY 735 GAACCTTAACAGTTTCCTTAAGAGCGCTGATATTTTCATCAATGGGTCCGACTTGAAGAA 794
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 194 -----ThrIleAlaAspArgValLeuGlyPro---GlySer 206

QY 795 GATTATATTCATTCATTAATCTTCAAGAAAGTCCACCTTCAAAATATGCAAACTCAT 854
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 207 PheTyrTrpGln---GlyGlnLeuLeuSerAspGlnValaGluIleIleSerLeuTyr 225

QY 855 GGCTCGTAGTGT-----GGTCACTCGCAAGAAAGCGGCTGAAG 896
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 226 AspProAsnValTyrSerIleLeuTyrAsnGlnIleValaIleThrArgThrAlaGlnAla 245

QY 897 CTGAAGACCTTCTGAAAGCTGTGAGAGATGATTTGATTCATTCAGTCAATGCACTAC 956
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 246 Ile-----PheAspAspSer 250

QY 957 TATTGAAATGACGAGCTGTACCAAGGAAATTTTCTTAACCTGATGTATGACATTT 1016
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 251 TyrLeu---GlyTyrSerValaIleValaGlyAspPhe---AsnGlyAspGlyIleGluAsp 268

QY 1017 TTTATTTGATCTGTCCAAAGATTTTCCAGGCTGTGAGAGAC-----AGGCTGGC 1070
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 269 PheValSerGlyValProArgAlaIleArgThrLeuGlyMetValTyrIleTyrAspGly 288

QY 1071 AAGAAGTTC-----TGTTAGGAGAGAAACAGCTCTGATATGAGGC--- 1112
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 289 LysAsnMetSerSerLeuIleAsnPheThrGlyGlnGlnMetAlaIleTyrPheGly 308

QY 1113 -----GAGCGGCC 1121
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 309 SerValaIleAlaThrAspIleAsnGlyAspAspTyrAlaAspValPheIleGlyAlaPro 328

QY 1122 TTGCTATCCGACACTTTGACGCTGGC---TTTATGCTGGCGATTAATTTGGCGTCA 1178
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 329 LeuPheMetAspArgGlySerAspGlyLeuGlnGluValGlyGlnValaSerValSer 348

QY 1179 GCCCGAATG-----GGAATGAAGTGTGTAGT 1205
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 349 LeuGlnArgAlaValaGlyAspPheGlnThrThrIleLeuAsnGlyPheGluValaPheAla 368

QY 1206 AGCGAAGTATTTCTTTGAGACAGAAATCAATTTAGTGAATGAATCTTGATCTTTTA 1265
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 369 Arg-----PheGlySerAlaIleAlaProLeuGlyAspLeuAspGlnAspGlyPhe 385

QY 1266 CCTGATTTATGGCTATCTCTTCTGTTCAGAAATTTGGTGGCAGCAAGTGTATGTGCA 1325
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 386 AsnAsp-----IleAlaIleAla 391

QY 1326 AGCGTCAAGGTTCAAGAGAAAGAGCTTGAATACCTTCAT----- 1370
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 392 AlaProTyrGlyGlyGlnAspIleGlyLeuValIlePheAsnGlyArgSerThr 411

QY 1371 TGCACAAACTACTGACAAATCAAGTATTAAGAGAGATTTAACTCTG----- 1418
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 412 GlyLeuAsnSerValProSerGlnIleLeuGlnGlyGlnIlePheAlaGlnIleSerMetPro 431

QY 1419 -----TATGCATTAACCTCATTAACCTCAACAACTCTGCGTTACCTTCTCT--- 1469
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 432 ProSerPheGlyTyrSerMetLeuGlyAlaThrAspValaAspArgAsnGlyTyrTrpAsp 451

QY 1470 -----TTTCTTAACAGAGAGTGAATATGATCTTTAAGACCTTTGAGA 1514
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 452 LeuValaIleGlyAlaPheGlyValaAspArgAlaValaIleuTyrArgAlaArgProVal-- 470

QY 1515 CCTCATGATTTAATTTCAAATCTGTCAACTCAAT---GGTCTTAAGCTTAAG----- 1565
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 471 -----ValThrValAsnAlaGlyLeuGlnValaTyrProSer 482

QY 1566 ---ATGCTGATGATCAACCTTGGCACTTTATATGAAAGAAACCTCTCGCGCGAGAGT 1622
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 483 IleLeuAsnGlnAspAsnLysIleCysProLeu-----ProGlyThr 496

QY 1623 TCACGTGGGCTTGCAGCTTTCATATATAGTTTGTGTGATGAAGAAATGCCAA 1676
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
  
```





Db 1943 laseglnleuphetrrtyrleuphenaenphelyalala 1957

RESULT 7

655765  
chitinase (EC 3.2.1.14) precursor - Aeromonas sp. (strain No. 10S-24) (fragment)  
C:Species: Aeromonas sp.  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #ext\_change 03-Nov-2000  
C:Accession: 655765  
R:Shiro, M.; Ueda, M.; Kawaguchi, T.; Arai, M.  
Biochim. Biophys. Acta 1305, 44-48, 1996  
A:Title: Cloning of a cluster of chitinase genes from Aeromonas sp. No. 10S-24.  
A:Reference number: 655762; MUID:96180984; PMID:6605248  
A:Accession: 655765  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-287 <SH1>  
C:Cross-references: GB:D63139; EMBL:D63142; NID:9667986; PIDD:BA09629.1; PID:dl010273;  
C:Keywords: glycosylase; hydrolase

Alignment Scores:

Pred. No.:	1-36	Length:	287
Score:	113.50	Matches:	51
Percent Similarity:	32.11%	Conservative:	10
Best Local Similarity:	26.84%	Mismatches:	65
Query Match:	3.64%	Indels:	64
DB:	2	Gaps:	7

US-10-676-079-3 (1-1721) x 655765 (1-287)

QY 89 GCCCGCCGCGGTATGATGCTGCTGCTGCGGCC-----GCTGGTCCCTGCTC 136

Db 53 AlAlAlValThnHtHrAlatGValGtAlaGtYtTpAnPrSeRhtPrOthr 72

QY 137 CCCGCGCC-----CTGCGCCGACCTGC 160

Db 73 LeutPrGlnArLeuGtGlyValArThrGlySerAlatHrPrOthrPrOglPrOthAla 92

QY 161 GCAGCAGCAGCAGCTGCGTGCAGCTGACCTTCTTCAACCCAGAGCCGCTGCTGAG 220

Db 93 AlArPrOthAlatHrPrOlysPro--ThrThrThrProAlaPrOthValThr--ProAl 111

QY 221 CCCCTGTTCTGTGCGCGTGCACATTG-----ACGCCAACCTGGCCACGGA 265

Db 111 lArPrOthMeCysPrOSeRhtatHrPrOlysPrOthSerThrPrOValTtPrOthr 130

QY 266 CCCCGGTTCTGATCTT----- 283

Db 131 PrOAlatHrPrOthrPrOglYValArGtYValAlLeuLeuPrOAlATPrAnThrGly 150

QY 284 -----CTGGGTTCTGCAAGCTTGATCTTGCCTTGGCCAGAGCTGTGCTCGTA 334

Db 151 ThrAlatYrGlnGtGlyValGlnValSerYrGlnGtYrGlnAla----- 168

QY 335 CCTGAGTTGGTGTGCACCAAGACAGACTTCTAAATTTCGATGCCAAGAGGAATCAAC 394

Db 169 -----LysTrPrTPrThrGlnGtYrGlnAlaPrO----- 177

QY 395 CTTGAAAGAGAGAACTTACTGGCAATCTGAAGTCAACAGATATTGGCAATATGATC 454

Db 178 -----SerSerSerThrGlyArPrGlnPrOthPrThr 188

QY 455 CATCCCTCTGATGTGAGAGAGAGAGTGAAGTTAGGTTGAAATGACCCCTACAGAGCAATTGCT 514

Db 189 ArPrAnGtYValCysGtYrPrOthPrOthPrOglYrPrOAlatHrPrOthPrOValAla 208

QY 515 ACTCCGAGAACTACACAGAAAA 538

Db 209 ThrAlatHrPrOlysPrOthPrGt 216

RESULT 8

IC5273

paired type homeobox protein, NBP - human  
C.Species: Homo sapiens (man)  
C.Date: 16-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 24-Sep-1999  
C.Accession: J05273  
R.Yokoyama, M.; Niehl, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.  
DNA Res. 3, 311-320, 1996  
A.Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific cDNAs  
A.Reference number: J05272; MUID:97191543; PMID:9039501  
A.Contents: neuroblastoma cell  
A.Accession: J05273  
A.Status: nucleic acid sequence not shown  
A.Molecule type: mRNA

A.Residues: 1-314 <YOK>  
A.Cross-references: DDBJ:D82344; NID:g1841337; PIDN:BA11555.1; PID:d1012222; PID:g1841337  
C.Comment: This protein is a transcriptional repressor involved in regulating gene expression  
C.Superfamily: unassigned homeobox proteins; homeobox homology  
C.Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F\_99-155/Domain: homeobox homology <HOX>

---

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1.66	314	34			
Percent Similarity:	112.50					
Best Local Similarity:	46.43*			5	20	
Query Match:	3.73*				25	
DB:	2					4

US-10-676-079-3 (1-1721) x J05273 (1-314)

```

Oy      231 AACCCAGAGATGATGAGAACCGCGGGTCCGTGGCCAGTTGGCTCATGTGACGACA   232
Db      211 Aaaglyglyglyglyly-----glyProSerProAla-----          221

Oy      231 GGAAcGAGGGGCTCAcAGGTCAGCGCGCTCTGCTGTAAGAATCAGGTCcACGAGT   172
Db      222 -----glyAlaProglyAlaAlaAglyProgly-----glyPro-----  232

Oy      171 CCTGCTTGCCGACAGTGCAGGGGCCAGCGCCAGGGAGAGGGAGACCAGCGCCCCAGGA  112
Db      233 -----                    -glyglygluProglyLygelylAlaAlaAlaAla  245

Oy      111 GCAGCAGATcAGCGCGCGCGCGCGAGCGcAGCGCTTCAGAcCGCACAGATCTGGGCTcAC  52
Db      246 AlaaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla  265

Oy      51 CTGGCTGCTCCC  40
Db      266 AlaGlyglyPro  269

RESULT 9
Fe4383
Hypotheoretical protein MJ0670 - Methanococcus jannaschii
C.Species: Methanococcus jannaschii
C.Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C.Accession: Fe4383
R.Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, C.E.; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A.Authors: Kalne, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Woese, C
A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A.Reference number: Ae4300; MUID:96337999; PMID:8688087
A.Accession: Fe4383
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-356 <BUL>
A.Cross-references: GB:U67514; GB:L77117; NID:g2826304; PIDN:AAB98664.1; PID:g1591384; T
C.Genetics:
A.Map position: REV5965956-595886
A.Start codon: GTG

Alignment Scores:
Pred. No.: 1.71 Length: 356

```



```

QY 648 GATTTCAGTGGACAGTTCTAAT-----GCTCAGTTCCTC 683
DB 241 GtuleuileIleThrThrSerLysasnleuValPheSerSerileSerSerAlaIleuArg 260
QY 684 CTGGAC-----TACTGCTCTCCAAAGGGTAT----- 710
DB 261 leuAsnAspLeuGlnAsnMetPheSerAlaPheArgProTyrGlyGlnGlnAlaIleuVal 280
QY 711 -----AACATTTCTGGGAAGTACAGCAATGAACCTTAACAGTTTC 749
DB 281 GlnAspPheAlaHisIleArgSerleuLysTrpPheSerAsnAspValGlnIleuSerleu 300
QY 750 CTTAAGAAAGCT-----GATATTTTC-----ATCAATGGCTCGCAGTTAGAA-- 791
DB 301 IleArgAlaIleuIlePheAsnAspMetPheProTyrPheAsnLysGlnIleuValAspThr 320
QY 792 ---GAAGATTATTTCAATTGCAATAACTTTAAGAAAG----- 827
DB 321 LysAlaAspGlyIlePhePheLeuArgLeuLeuArgLysAsnPheGlnHisIleAsn 340
QY 828 -----TCCACCTTCAAAAT----- 842
DB 341 AspValLysAspPheHisIleGlnValIleLysTyrLeuAsnSerGlnPheLysAsnAsn 360
QY 843 ---GCMAAATCTATGCTCTGATGTTGCTCAGCTCGAAGAAGACGGCTAAGATGCTG 899
DB 361 TyrSerThrIleuMetThrSerSerLysThrGlnAspArgLysSerHisAsnMetPro 380
QY 900 AAGAGCTTCTGAAGGCTGCTGAGACAGTG-----ATTGATTCAGTT 941
DB 381 SerSerIleLeuAspAspGlyAsnLysIleGlyMetHisValSerProIleAspGlu-- 399
QY 942 ACATGGACCTACATTCATTGAATGA-----CGG 971
DB 400 ---TyrSerHisPheIleAspAsnAspGluProLeuTTrpArgAspLysValIleProLys 418
QY 972 ACTGCTACCAAGGAAGATTTTAAACCTGATGATTTGATTTTATTCATCTGTG 1031
DB 419 IleTyrThrAsnGlnIleThrProThrProAspAlaSerAlaIlePheAspSer----- 436
QY 1032 CAAAAGTTTCCAGGTGTTGAGACACAGC-----CTGGCAAGAGAGTCTGG 1082
DB 437 HisLysIleTyrAlaIleIleSerLeuLeuArgTyrTyrLeuProGlnLysArgLysPhe 456
QY 1083 TTAGAGAAACAAGCTCTGCATATGAGAGCGGAGCGCCCTTGCTATCCGACACCTTGA 1142
DB 457 -----PheArg 458
QY 1143 GCTGCTTATGTCGTGATTAATGGCTGTCAGCCCGAATGGAATAGAGTGTG 1202
DB 459 IleTyrTyrLeu-----ProSerIlePhe 466
QY 1203 ATGAGCAAGATTTCTTTGAGCA-----GAAACTACATTAGTGATGAAAACTTC 1256
DB 467 LysAlaGlyIleLeuTyrTyrIleLysPheAlaGlnIleuTyrPheMetGlnGlyCysLeu 486
QY 1257 GATCCTTATCCGATATTTGCTATCTCTGTTCAAGAAATTTGGTGGC-----ACC 1310
DB 487 GlnArgLeuValIleGlnSerleuGlnIleuGlnProSerLeuValHisAlaIleAsn 506
QY 1311 AAGGTGTTAATGCAAGCTGCAAGGTTCAAGAGAAAGAACTTCAGATATACCTTCAT 1370
DB 507 AsnLeuIleLysSerSerIleGlnSerleuLys-----Asn 518
QY 1371 TGCACAAACATGACATCAAGTAAAGATTAAGAGATTTAATCTGTATGCCATTAAC 1430
DB 519 ValThrValThrSerAspAspLysThrSerSerGlyValIleIleLeu----- 534
QY 1431 CTCCAATAGGTACCAAGACTTGGGTTACCTTATCCCTTTTCAACAAGAGATGAT 1490
DB 535 -----SerTyrLysGlnPheLys-----SerLeuSerGlnValAsn 546
QY 1491 AATACTTCTTAAGACCTTTGGAGACTCATGATTACTTTCCAAATCTGTCCAACATCAT 1550

```

```

DB 547 LysAspPheAsnGlnuProPheTrpProAsnGlnSerIleAlaAsnSer----- 562
QY 1551 GGTCTAATCTTAAGATGTGTGATGATCAAACTTGGACCTTTAATGGAAGAACTCTC 1610
DB 563 -----TrpProAspPheAlaAsnLysGlnIleu 571
QY 1611 CGGCAGAGAGTTCACGTGGCTTCGACCTTTCATTAAGTTTCTGTATAGAAAT 1670
DB 572 LysArgGlyGlnIleLeu-----GlnAspAlaPheAlaPheHisLeuPheGlnIle-GluLe 590
QY 1671 GCCAAAGTTCCTGCTTCATCTGAAAT 1698
DB 590 uProIleIleIleAspThrThrArgAsn 599

RESULT 11
A27421
Integrin alpha-5 chain precursor - human
N:Alternate names: CD51 antigen; serum spreading factor; vitronectin receptor alpha chain
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1989 #sequence revision 31-Mar-1989 #text change 31-Dec-2000
C:Accession: A27421; S47541; B29418; A26482; A32287; B32287; A30298; A35035
R:Suzuki, S.; Argraves, W.S.; Arai, H.; Languiño, L.R.; Pierschbacher, M.D.; Ruoslahti, I
J. Biol. Chem. 262, 14080-14085, 1987
A:Title: Amino acid sequence of the vitronectin receptor alpha subunit and comparative ex
A:Reference number: A27421; MUID:88007656; PMID:2443500
A:Accession: A27421
A:Molecule type: mRNA
A:Residues: 1-1048 <SUZ>
A:Cross-references: GB:M14648; GB:J02826; GB:M18365; NID:9340306; PIDN:AAA36808.1; PID:93
R:Donahue, J.P.; Sugr, N.; Hawiger, J.
Biochim. Biophys. Acta 1219, 228-232, 1994
A:Title: The integrin alpha(v) gene: identification and characterization of the promoter
A:Reference number: S47541; MUID:94368864; PMID:7522056
A:Accession: S47541
A:Molecule type: DNA
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-61 <DON>
A:Cross-references: EMBL:U07375
A:Title: Comparison of cDNA-derived protein sequences of the human fibronectin and vitron
R:Fitzgerald, L.A.; Poncz, M.; Steiner, B.; Rall Jr., S.C.; Bennett, J.S.; Phillips, D.R.
Biochemistry 26, 8158-8165, 1987
A:Title: Comparison of cDNA-derived protein sequences of the human fibronectin and vitron
A:Reference number: A30526; MUID:88163472; PMID:2450560
A:Accession: B29418
A:Molecule type: mRNA
A:Residues: 1-433 <FIT>
R:Suzuki, S.; Argraves, W.S.; Pytela, R.; Arai, H.; Krusius, T.; Pierschbacher, M.D.; Ruc
Proc. Natl. Acad. Sci. U.S.A. 83, 8614-8618, 1986
A:Title: cDNA and amino acid sequences of the cell adhesion protein receptor recognizing
A:Reference number: A26482; MUID:87041504; PMID:2430295
A:Accession: A26482
A:Molecule type: mRNA
A:Residues: 413-1048 <SUZ>
R:Cheresh, D.A.; Smith, J.W.; Cooper, H.M.; Quaranta, V.
Cell 57, 59-69, 1989
A:Title: A novel vitronectin receptor integrin (alpha-v beta-x) is responsible for discri
A:Reference number: A32287; MUID:89195223; PMID:2467745
A:Accession: A32287
A:Status: preliminary
A:Molecule type: protein
A:Residues: 31-35, 'X', 37-41 <CHR>
A:Experimental source: melanoma cell M21
A:Accession: B32287
A:Status: preliminary
A:Molecule type: protein
A:Residues: 31-33, 'X', 35, 'X', 37-41 <CH2>
A:Experimental source: lung carcinoma cell UCLA-P3
R:Iliam, S.C.T.; Plov, E.F.; D'Souza, S.E.; Charest, D.A.; Prellinger III, A.L.; Ginsberg, M
J. Biol. Chem. 264, 3742-3749, 1989
A:Title: Isolation and characterization of a platelet membrane protein related to the vit
A:Reference number: A30298; MUID:89139425; PMID:2465293
A:Accession: A30298
A:Molecule type: protein

```



A:Experimental source: cv. Fribo, seed coat

C:Genetics:

A:Gene: CWINV1

C:Superfamily: beta-fructofuranosidase

C:Keywords: cell wall; glycoprotein; glycosidase; hydrolase

#### Alignment Scores:

Pred. No.:	2,32	Length:	575
Score:	111.50	Matches:	72
Percent Similarity:	35.71%	Conservative:	48
Best Local Similarity:	21.43%	Mismatches:	107
Query Match:	3.57%	Indels:	109
DB:	2	Gaps:	19

US-10-676-079-3 (1-1721) x T12094 (1-575)

```

QY 198 CAGGAGCCGCTGCACCTGTGAGC-----CCCTGTTCTGTCC 236
DB 228 LYSHPSPDLEHISERLALYSARGTHRGlyMETTRGlyUCYSPROPHETYP 247
QY 237 GTCACCATGAGCCCACTGGCCAGGACCCGGCTTCTATCTCTGGGTCTCA 296
DB 248 VALSERLEUGLUGLYLSANGLYLEUASP-----LEUSERMETMETGLYSN 265
QY 297 AAGCTTCGACCTGGCCAGGAGCTGTCT-----CCT 329
DB 266 VALLYSHIVALLLEULYSANSESERLEUAPRIETHRARGTYRGLUTYRTHRI 285
QY 330 GCGTACCTGAGG-----TTTGGT 347
DB 286 THRTYRLEUGLHANSGLNAPRYTYRIEPROAPRYTHRSEGLUAPRYTRGly 305
QY 348 GGCACCAACAGACACTTC-----CTAATTTGCATCCCAAGAG 386
DB 306 GLYLEUARGTYRSPRYGLYSNPHERYALASERYSESPHEHAPROTHRYLS 325
QY 387 GATCAACCTTGAAGAGATTACTGG-----CAATCTCAAGTCAACCG 434
DB 326 -----ASNARGARGILEIETTRPGlyTRPALASNGLSERAPHRYSGLU 341
QY 435 GATATTGGCAATATGGA-----TCCATCCCTCGTAGTG----- 470
DB 342 ASPAPRYVALLYSGLYTRPALAGLYLEGINALALEPROARGHVALTRPLEUAP 361
QY 471 -----GAGAGAAAGTTAGGTTGAAATGCCCTTACAGAGCAATTGCTAC 521
DB 362 SERSEARGARGINLEUARG--GINTPRPROVALGLUGLHLEUASNARGRYGLY 380
QY 522 -----GACACTACCAAAAAGTTCAAGAACAGACCACTCAAGAGCTCTGA 572
DB 381 LYSGLINVALGLIMELYSANRGLYLSULYLSGLYLYR-----LEU 396
QY 573 GATGTCCTATACACTTTGCAAACTGCTGAGACTGAGACTTATCTTTGGCTTA 632
DB 397 GLUVALLYGLYILETHRALASERGINALAPRYVALGLUVALTHRPHESER 416
QY 633 TTAATTAAGACAGACATTGGACGTGAAACAGTTCTAATGCTCAAGTCTCTG 692
DB 417 LEUAPRYVALGLUVALPHEASPRPROANTTRPGlyUASNALAGLU----- 433
QY 693 TGCCTTTCAGAGGGGATATACTTTCTTGGAACCTAGCAATGAACCTAAGTT 752
DB 434 CYSHALGGLNLYSGLYSERLYSVALARGLYVALGly-----PROHEGLYLE 451
QY 753 AAGAGGCTGATATTTTCATCAATGGGTGCAATTAGGAAGATATTAATTCATT 809
DB 452 THRLEUALA-----SERLYSLEUGLUGLUTYRTHRSEVALPHE 465
QY 810 -----CATTAACCTCTA-----AGAAAGTCC 830
DB 466 PHEARGVALPHELYSVALALASNLYSHLYSILEUWECYSESERAPALYSER 485
QY 831 ACCTTCAAAAATGCAAACTGATGTCT-----GATGTT 866

```

```

DB 486 SERSELEUASNARGGLUTEUTRYRSPROSERPHEALAGLYPHEVALASNVALP 505
QY 867 GGTCAAGCTCCGAAAGAAACCGCTTAAGTCTGAAGAGCTTCTGAGAGCTG 926
DB 506 GLYASNANLYSGLYLSERLEUARGSERLEU----- 516
QY 927 GTGATTGATTCAGTTACATGAGATCATCACTAATTGAATGAGCGACT 974
DB 517 ---ILEASPHLSERVALVALGLUSERPHEGLYVALGlyLYSFRH 531

RESULT 13
T13425
regulatory protein K10, oocyte-specific - fruit fly (Drosophila melanogaster)
N:Alternative names: protein Bg:30B8.5
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
C:Accession: T13425; A28826
R:Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17668
A:Accession: T13425
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-463 <MUR>
A:Cross-references: EMBL:AL009195; NID:e135203; PID:e1202207; PIDN:CA15702.1
R:Prost, E.; Deryckere, F.; Koos, C.; Haendlin, M.; Fanesco, V.; Mohler, E.
Genes Dev. 2, 891-900, 1988
A:Title: Role of the oocyte nucleus in determination of the dorsoventral polarity of Dro
A:Reference number: A28826
A:Accession: A28826
A:Molecule type: mRNA
A:Residues: 1-276, 'HH', 279-281, 'VDHHR', 287-463 <PRO>
A:Cross-references: GB:X12836; NID:98148; PID:g295771
C:Genetics:
A:Gene: FlyBase:fs(1)K10
A:Cross-references: FlyBase:FBgn0000810
A:Map position: X
A:Introns: 432/3
C:Keywords: DNA binding

Alignment Scores:
Pred. No.: 2,39 Length: 463
Score: 111.00 Matches: 38
Percent Similarity: 42.86% Conservative: 10
Best Local Similarity: 33.93% Mismatches: 40
Query Match: 3.56% Indels: 24
DB: 2 Gaps: 4

US-10-676-079-3 (1-1721) x T13425 (1-463)
QY 21 CTGCGCGGACGCTGCGGGGAGAGCAGGAGCCCAAGATGCTGCGCTGAG 80
DB 215 ILEARGLYGLYPHEASNGLYPHEINARGLYPRO----- 227
QY 81 CTGCGCTGCGCGCGCGCTGATGCTGCTCTGCGGCGCGTGGTCTCCCTCC 140
DB 228 PROASNARGPROPRO-----ARGLEUWETGLYPROPOMETGLYPROMETGLY 246
QY 141 GGGCGCGTCCCGCCGACCTGCGCAACAGACAGACGTCGAGACCTTGATTCAC 200
DB 247 GLY--PRO-----ARGGLYPROGLYPROMETGLYPROG 257
QY 201 GAGCGCTGACCTGTGAGCCCTGCTCTGTCCTGACCATTAAGCCAACTGAC- 259
DB 257 LYGlyProTYRProGLINMETProPHEProProVALProGLYMETARGGLYProGLY 277
QY 260 -----CAGGACCCGCGGTTCTCATCTCTCTGCGGTCTCCAAAGCTTGT 305
DB 277 IOMETGLYPROMETGLYGLYProProProProProProProLeuPHEMETARGARG 297
QY 306 ACCTTGGCGAAGAGCTGTCTCTGCTGACCT 337

```

Db 297 lypProGlyProGlyPrometMetGlyValPro 307  
RESULT 14  
T10666  
hypothetical protein F6E21.40 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 16-Jul-1999 #sequence, revision 16-Jul-1999 #text change 16-Feb-2001  
R/Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft  
submitted to the Protein Sequence Database, June 1999  
A/Reference number: Z16533  
A/Accession: T10666  
A/Molecule type: DNA  
A/Residues: 1-670 <BEV>  
A/Cross-references: EMBL:AL049914; GSPDB:GN00062; ATSP:F6E21.40  
A/Experimental source: cultivar Columbia; BAC clone F6E21  
C/Genetic:  
A/Map position: 4  
A/Intons: 47/3; 87/1; 123/3; 203/3; 230/2; 255/3; 284/3; 305/1; 335/3; 347/3; 370/3; 39  
C/Superfamily: Schizosaccharomyces pombe negative regulator of mitosis skb1  
Alignment Scores:  
Pred. No.: 2.64 Length: 670  
Score: 111.00 Matches: 123  
Percent Similarity: 36.36% Conservative: 77  
Best Local Similarity: 22.36% Mismatches: 194  
Query Match: 3.56% Indels: 156  
Gaps: 33  
US-10-676-079-3 (1-1721) x T10666 (1-670)  
QY 213 CTGATGAGCCCTGTCCTGTCCTGTCACCATTCAGCCCACTGGCCAGCCCGG 272  
Db 47 LeuValAspProSerTyrArgProSerLeuValGluValAsn--GlyValAspThrGln 65  
QY 273 TTCTCATCTCTCGGGTTCTCCAAAGCTTGATCTGGCCAGAGCTTGCTCTGGC 332  
Db 66 ValLeuProValCysGlySerAspLeuVal-----LeuSerProSer 79  
QY 333 TAC-----CTGAGGTTTGTCGACCAAGACAGCTTCGAAATTTTCATCCCAAG-- 383  
Db 80 GluTrpSerSerHisValValGlyLysIleSerSerTrpIleAspLeuAspSerGluAsp 99  
QY 384 -----AAGAAATCAACTTTGAGAGAGAAATTACTGCAATCTCAA 425  
Db 100 GluValLeuArgMetAspSerGluThrThrLeuLysGluIleAlaTrpAlaThrHis 119  
QY 426 GTCAACACGAGATTTTGCAAATATGATCCATCCCTCTGATGTGGAGAGAAATTACGG 485  
Db 120 LeuSerLeuGluMetCysGlu-----ProAsp----- 128  
QY 486 TTGGAATGGCCCTACACAGAGCAATTGCTATCCGAGACATCCAGAAAAGTTCAAG 545  
Db 129 -----LeuThrArgProHisTyrLeuAlaGlyGlyLeu 139  
QY 546 AACAGCACTACTCAAGAGCTGTGA-----GATGTCTTAAC----- 584  
Db 140 ArgValSerCysGlySerSerPheIleSerAspGluThrPheLeuTyrLysIleThr 159  
QY 585 -----ACTTTTGCAACTGCTCA-----GAGCTGCACTTGATCTTTGGC 623  
Db 160 PheAsnGlnAlaLeuThrPheCysGlySerSerLeuPheCysLeuAsnValIleSerAla 179  
QY 624 CTAAATGCGTTATTAGACAGAGATTTGCAGAGGAAAGTTGATGCTCAAGTGTCTC 683  
Db 180 LeuLeuLeuTrpLeuValGlyValProLeuValLys-----SerGluGlyAspSerMet 196  
QY 684 CTGAGACTACTCTCTTCCAAAGGGGTATACATTTCTTGGAACTAGGCAAT----- 724  
Db 197 AspAsp-----ThrSerGluGlyLeuAsnAspSerTrpGluLeuLeuTrpAsnSerPheArg 214

QY 735 -----GAACCTAACGTTTCTTAAGAGCT--GATATTTTCATCAT----- 776  
Db 215 LeuLeuCysGluHisIleAspSerLysIleSerValAlaLeuAspValLeuSerThrLeuPro 234  
QY 777 -----GGTCGCGAGTTAGAGAAATTAATTAATTCATTAAGTCTTA 821  
Db 235 SerGluThrSerLeuGlyArgTrpMetCysGlyLysSer--ValArgAlaAlaIleLeuSer 253  
QY 822 AGAAAGTCCACTTCAAAATATGCAAACTCTATGCTCGATGTTGGACGCT----- 875  
Db 254 ThrAspAlaPheLeuThrAsnAlaArg-----GlyTyrProCysLeu 267  
QY 876 CGAAGAAAGCGCTTAAGTCTGTAAGAGACTTCCTG--AAGCTGTGTGAGAAAGCAT 932  
Db 268 SerLysArgHisGlnLysLeuIleAlaGlyPhePheAspHisAlaAlaGlnValValIle 287  
QY 933 GATTCAGTTACATGGCAGCTACTATTGTAATGACGGAGCTGCTACAGGGAAGATTTT 992  
Db 288 CysGlyLysProValHisAsnLeuGlnLysProLeuAspSerSerSerGluLysThrGlu 307  
QY 993 CTAAACCTGATGTATGACATTTTATTTTATCTGTCGCAAAAGTTTTCAGGTGTT 1052  
Db 308 LysAsnPro-----LeuArgIleTyrLeuAspTyrValAlaTyrLeuPheGlnLysMet 325  
QY 1053 GAGAGCACAGCGCTGCGCAAGAAAGTCTGTTAGAGAAACAAAGCTCTGATGTGAGAGC 1112  
Db 326 GluSerLeuSerGluGlnGluIleArgIleGluLysTyrArgAspPheLeuGlnAlaPro 345  
QY 1113 GGAGCGCCCTGTCCTATCCGACACCTTGACAGCTGCTTATGTCGTGATTAATTTGGCG 1172  
Db 346 LeuGlnProLeuMetAspAsnLeuGluAlaGlnThrTyrGluThrPheGlu----- 362  
QY 1173 CTGTGACCCGAAATGGAAATGAGATGTGATGAGGCAAGTATCTTTGAGAGAGAAAC 1232  
Db 363 -----ArgAspSerValLysTyrIleGlnTyrGln-----ArgAlaValGlu 376  
QY 1233 TACCATTTAGTGATGATGAATTCCTGATCCTTAACCTGATTAATGCTATCTCTGTTG 1292  
Db 377 LysAlaLeuValAspArg-----ValProAspGluLysAlaSerGluLeu----- 391  
QY 1293 AAGAAATTTGTTGGCCCAAGGTTTATG----- 1322  
Db 392 -----ThrThrValLeuMetValValGlyAlaGlyArgGlyProLeuVal 406  
QY 1323 --GCAAGCGTGCAGGTTCAAGAGA-----AGGAACCTTGAGTATACCTTCATTGC 1373  
Db 407 ArgAlaSerLeuGlnAlaAlaGluGluThrAspArgLysLeuLysValTyr----- 423  
QY 1374 ACAAAACATGACAATCCAAAGTATTAAGAAGATTTAACTGTATGCCATTAACCTC 1433  
Db 424 AlaValGluLysAsnProAsn-----AlaValValThrLeu 435  
QY 1434 CATACGTCACCAAG-----TACTTGGCGTTAACCTCAT 1466  
Db 436 HisAsnLeuValLysMetGluGlyTyrGluAspValValThrIleIleSerCysAspMet 455  
QY 1467 CCTTTTCTAAC-----AAGCAAGTGAATAATCTTCTAAGACCTTTGGAGCTCAT 1520  
Db 456 ArgPheTrpAsnAlaProGluGlnAlaIleAlaIleValSerGluLeuLeuGlySerPhe 475  
QY 1521 GGAATTCTTCCAAATCTGTCCACTCAATGCTTACTCTAAAGTGTGATGATCAA 1580  
Db 476 Gly-----AspAsnGlu 479  
QY 1581 ACCTGCGACCTTTAATG-----GAAAAACCTTCGCGCAGAGATTCAGGGGC 1631  
Db 480 LeuSerProGluCysLeuAspGlyAlaGlnArgPheLeuLysPro--AspGlyIleSer 498  
QY 1632 TTGCGAGCTTTCTCATATATGTTTTTTGTG 1661  
Db 499 IleProSer--SerTyrThrSerPheIle 507  
RESULT 15

## LURB11

annexin XI - rabbit

N.Alternate names: calyocyclin-associated annexin protein CAP-50

C.Species: Oryctolagus cuniculus (domestic rabbit) CAP-50

C.Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #ext\_change 16-Jun-2000

C.Accession: JH0694; PH0950; A38250; PS0263

R.Tokumitsu, H.; Mizutani, A.; Muramatsu, M.; Yokota, T.; Arai, K.; Hidaka, H.

Biochem. Biophys. Res. Commun. 186, 1227-1235, 1992

A.Title: Molecular cloning of rabbit CAP-50, a calyocyclin-associated annexin protein.

A.Reference number: JH0694; MUID:92378579; PMID:1380798

A.Accession: JH0694

A.Molecule type: mRNA

A.Residues: 1-503 &lt;TOX&gt;

A.Cross-references: DDBJ:D10883; NID:g471147; PIDN:BA01705.1; PID:g471148

A.Experimental source: Lung

A.Accession: PH0950

A.Molecule type: protein

A.Residues: 104-141;213-231;254-262;270-280;285-309;319-337;429-448;478-492 &lt;TK2&gt;

R.Tokumitsu, H.; Mizutani, A.; Minami, H.; Kobayashi, R.; Hidaka, H.

J. Biol. Chem. 267, 8919-8924, 1992

A.Title: A calyocyclin-associated protein is a newly identified member of the Ca(2+)/phosph

A.Reference number: A38250; MUID:92250478; PMID:1533622

A.Accession: A38250

A.Molecule type: protein

A.Residues: 104-141;213-223,'X',225-231;254-263;271-280;285-291,'X',293-300,'X',302-309;

C.Comment: This protein binds specifically to calyocyclin in a Ca2+ dependent manner.

C.Superfamily: annexin VII; annexin repeat homology

C.Keywords: calcium binding; duplication; endonexin fold; glycoprotein; phospholipid bin

F.201-272/Domain: annexin repeat homology &lt;AX1&gt;

F.212-228/Region: endonexin fold #status predicted

F.273-344/Domain: annexin repeat homology &lt;AX2&gt;

F.284-300/Region: endonexin fold #status predicted

F.356-428/Domain: annexin repeat homology &lt;AX3&gt;

F.368-384/Region: endonexin fold #status predicted

F.432-503/Domain: annexin repeat homology &lt;AX4&gt;

F.443-459/Region: endonexin fold #status predicted

F.58/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Alignment Scores:

Pred. No.:	2.67	Length:	503
Score:	110.50	Matches:	77
Percent Similarity:	36.41%	Conservative:	53
Best Local Similarity:	21.57%	Mismatches:	123
Query Match:	3.54%	Indels:	104
DB:	1	Gaps:	17

US-10-676-079-3 (1-1721) x LURB11 (1-503)

```

QY 110 GCTCTGGGGCGGCTGGTCCCTCTCCGCC-----TGGCGCCCTGCCCCGACC 157
    |||||
DB 76 AIAFProGlyGlyGlyTyrProProValProProGlyGlyPheGlyGlnProProThr 95
    |||||
QY 158 TCGCAGACGACGACGCTGTGA-----CCTGA----- 187
    |||||
DB 96 GlnProSerValProProTyrGlyValTyrProProGlyGlyAsnProProSergly 115
    |||||
QY 188 -----CTTCTTACCAGAGCGCGCTGCACCTGTGAGCGCCCTGCTCTGTC 235
    |||||
DB 116 ValProSerTyrProProPheProGlyAlaProValProGlyGlnProMetProPro 134
    |||||
QY 236 CGTCACCATTTAGCGCAACCTGGACGAGACCGCGGTTCCATCCTCTGGGTTCTCC 295
    |||||
DB 135 -----ProGlyHisGlnProProGlyProTyr---ProGlyGlnLeu 147
    |||||
QY 296 AAAGCTTCGACTTGGCCAGAGGCTTGTCTCTGCGTACCTGAGGTTTGTGGACCAA 355
    |||||
DB 148 ProValThrTyrProGlyGln-----SerProValProProPro-----GlyGln 162
    |||||
QY 356 GACACCTTCTTAATTTTGATCCCAAGAGAAATCAACCTTTGAAAGAGAAATTACTG 415
    |||||
DB 163 GlnProMetProSerTyr----- 168
    |||||
QY 416 GCAATCTCAAGTCAACGACGATATTGCAATATGATCATCCCTCTGATGTGAGAGA 475
    |||||

```

```

DB 169 -----ProGlyTyr----- 171
    |||||
QY 476 GAAGTTACGTTGGAATGCCCTTACCGAGAGCA---ATTGCTACTCCGAGAACTACCA 532
    |||||
DB 172 -----ProGlySerGlyThrValThrProAlaValProPro 183
    |||||
QY 533 GAAAAGTTCAAGAACACGACCTACTCAAGAAAGCTGTAGATGTCATACCTTTTGC 592
    |||||
DB 184 -ValGlnPheGlyAsnAlaGly-----ThrIleThr 193
    |||||
QY 593 AAATGCTCAGACACTGACCTTGATC----- 617
    |||||
DB 193 rAspAlaSerGlyPheAspProLeuArgAspAlaGluValLeuArgLysAlaMetLysGlu 213
    |||||
QY 618 -TTTGCGCTTAAT-----GGCTATTAGAACACGACATTGCGAGTGAACAGTTCTTA 670
    |||||
DB 213 yPheGlyThrAspGluGlnAlaIleIleAspCysLeuGlySerArgSerAsnLysGlnArg 233
    |||||
QY 671 TGCTCAGTTGCTCGTGACTACTGCTCTTCCAGGCGGTATTAACATTTCTTGGAACTAG 730
    |||||
DB 223 gGlnGlnIleLeuLeuSerPheLysThrAlaTyrGlyAspLeuIleLysAspLeuLys 253
    |||||
QY 731 CAATGAACCTTAACAGTTTCTTAAGAACGCTGATATTTCATTCATGAGTGCAGTATAG 790
    |||||
DB 253 sSerGlnLeuSerGlyAsnPheGlnLysThrIleLeuAlaLeuMetLysThrProIleLe 273
    |||||
QY 791 AGAAGATTATATTCATTCATTAACCTTCAAGAAATTCACCTTCAAAAATGCA----- 845
    |||||
DB 273 uPheAspAlaTyrGlnIleLysGlnAlaIleLysGlyLysThrAspGluAlaCysLe 293
    |||||
QY 846 ----AAACCTATATGCTCTGATGTGTGTCACGCTCGAAGAAAGCGGTAAAGATCTGAA 901
    |||||
DB 293 uIleGlnIleLeuAlaSerArgSerAsnGlnHisIleArgGlnLeuAsnLysAlaTyrLys 313
    |||||
QY 902 GAGC---TTCTGAGAGCTGTGAGAGATGATGATTCATTCATGATGATGATGATGATGAT 956
    |||||
DB 313 sThrGlnPheLysGlyThrLeuGlnGluAlaIleLeuArgSerAspThrSerGlyHisPheGlu 333
    |||||
QY 957 -----TATTTGAATGACGAGACTGCTACCAAGGAGATTTTCTAAACCTGATGATTT 1009
    |||||
DB 333 nArgLeuLeuIleSerLeuSerGlnGlyAsnArgAspLysThrAsnValAsp---Me 352
    |||||
QY 1010 GGACATTTTATTCATCTGTGCAAAAGTTTCCAGGTGATGAGAC 1058
    |||||
DB 352 tSerLeuValGlnArgAspValGlnGlnLeuTyrAlaAlaGlyGlnAsn 368
    |||||

```

Search completed: August 28, 2004, 08:01:43  
 Job time : 96.5 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: August 28, 2004, 07:42:36 ; Search time 36.5 Seconds  
(without alignments)  
4910.287 Million cell updates/sec

Title: US-10-676-079-3  
Perfect score: 3119  
Sequence: 1 ctgagcttcgactctcg.....atactagctgactgactg 1721

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 segs, 52070155 residues  
Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV-xlp  
-MODE=frame+.n2p.model -DEV-xlp  
-Q=/cgn2.1/USFTO.spool.p/US10676079/runat\_28082004\_084225\_23803/app\_query.fasta.1.1863  
-DB=SwissProt\_42 -QPM=faetan -SUFFIX=rbp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNIT=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10676079 @CNC 1.1.48 @runat\_28082004\_084225\_23803 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEOUTER -NEG\_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124.5	4.0	1044	1 ITAV_MOUSE	P43406 mus musculu
2	112.5	3.7	314	1 PKXB_HUMAN	O99453 homo sapien
3	112.5	3.7	314	1 PKXB_MOUSE	O35690 mus musculu
4	112.5	3.6	356	1 Y670_METUA	Q58084 methanococ
5	112	3.6	636	1 GIDA_LACPL	Q882x6 lacobacill
6	112	3.6	688	1 YB9F_YEAST	P38338 saccharomy
7	112	3.6	1048	1 ITAV_HUMAN	P06756 homo sapien
8	111	3.6	463	1 K10_DROME	P13468 drosophila
9	110.5	3.5	503	1 ANXB_RABIT	P33477 oryctolagus
10	110	3.5	607	1 THIC_AGRIS	O8ucv9 agrobacteri
11	109.5	3.5	505	1 ANXB_HUMAN	P50995 homo sapien
12	109	3.6	627	1 SPDA_NEPCL	P46804 nephila cla
13	109	3.5	694	1 RPD3_HUMAN	O92310 homo sapien
14	108	3.5	503	1 ANXB_MOUSE	P97384 mus musculu
15	108	3.5	641	1 SYM_CLOTE	O899d9 clostridium
16	107	3.4	1247	1 CTBD_MOUSE	O35927 mus musculu
17	106.5	3.4	664	1 AROB_FUSAN	O8f747 fusobacteri
18	106.5	3.4	1203	1 HCN4_HUMAN	O9yq44 homo sapien

19	106	3.4	646	1 THIC_VIBRA	Q87kf0 vibrio para
20	106	3.4	1579	1 SSR2_YEAST	P53599 saccharomyc
21	105.5	3.4	1031	1 TERT_EUPAE	O00939 euploes ae
22	105	3.4	670	1 SSR4_SCHPO	O42667 schizosach
23	105	3.4	1048	1 SRM4_RAT	O63627 rattus norv
24	105	3.4	1500	1 CPBW_HUMAN	P31327 homo sapien
25	104.5	3.4	716	1 RRP2_TAKIT	O91742 influenza a
26	104.5	3.4	788	1 PUR2_SCHPO	O91742 influenza a
27	104	3.3	432	1 PURA_AQUAE	O67321 aquifex aeo
28	103	3.3	796	1 PKR_CLOAB	O97je3 clostridium
29	103	3.3	1586	1 SN22_HUMAN	P51531 homo sapien
30	102	3.3	481	1 CB12_MOUSE	Q8K3m5 mus musculu
31	102	3.3	503	1 ANXB_BOVIN	P27214 bos taurus
32	102	3.4	1205	1 S122_MOUSE	P55012 mus musculu
33	102	3.3	1231	1 CPFH_HUMAN	P08603 homo sapien
34	102	3.3	1556	1 PROS_DROVI	O9ue61 drosophila
35	101.5	3.3	478	1 CB12_HUMAN	Q9cvt7 homo sapien
36	101.5	3.3	492	1 MM11_MOUSE	Q02853 mus musculu
37	101.5	3.3	495	1 P033_MOUSE	P31361 mus musculu
38	101.5	3.3	497	1 P033_RAT	O63262 rattus norv
39	101.5	3.4	1262	1 CA13_CHICK	P12105 gallus gall
40	101.5	3.3	1572	1 BA12_HUMAN	O60241 homo sapien
41	101	3.3	401	1 HB9_HUMAN	P50219 homo sapien
42	101	3.3	518	1 MTGO_MOUSE	Q60754 mus musculu
43	101	3.2	595	1 THIC_BACHD	Q9hb14 bacillus ha
44	101	3.2	644	1 SYM_CLOAB	O97ew5 clostridium
45	101	3.2	666	1 TKT_BACHD	Q9xad7 bacillus ha

#### ALIGNMENTS

RESULT 1  
ITAV\_MOUSE  
ID ITAV\_MOUSE STANDARD; PRT; 1044 AA.  
AC P43406;  
DT 01-NOV-1995 (rel. 32, Created)  
DT 01-NOV-1995 (rel. 32, Last sequence update)  
DT 28-FEB-2003 (rel. 41, Last annotation update)  
DE Integrin alpha-V precursor (Vitronectin receptor alpha subunit)  
DE (CD51 antigen).  
GN ITGAV.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1; TISSUE=Kidney;  
RX MEDLINE=96176309; PubMed=8601592;  
RA Wada J., Kumar A., Liu Z., Ruoslahti E., Reichardt L., Marvaldi J.,  
RT Kanwar Y.S.;  
RT "Cloning of mouse integrin alphaV cDNA and role of the alphaV-related  
RT matrix receptors in melanophytic development.";  
RL J. Cell Biol. 132:1161-1176(1996).  
CC -!- FUNCTION: The alpha-V integrins are receptors for vitronectin,  
CC cytoactin, fibronectin, fibrinogen, laminin, matrix  
CC metalloproteinase-2, osteopontin, osteomodulin, prothrombin,  
CC thrombospondin and von Willebrand factor. They recognize the  
CC sequence R-G-D in a wide array of ligands. Alpha-V integrins may  
CC play a role in embryo implantation, angiogenesis and wound  
CC healing. Mice expressing a null mutation of the alpha-V subunit  
CC gene survive until late in embryonic development and occasionally  
CC even to birth. They demonstrate cleft palate, and defective  
CC development of CNS and gastrointestinal blood vessels.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA  
CC DISULFIDE BOND. ALPHA-V ASSOCIATES WITH EITHER BETA-1, BETA-3,  
CC BETA-5, BETA-6 OR BETA-8.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR		EMBL; U14135; AAC52497.1; -;	
DR	PIR; T10050; T10050.		
DR	HSSP; P06756; I0V2.		
DR	GGO; MG1;96608; R'edav.		
DR	MG; GO:0001568; P:blood vessel development; IMP.		
DR	InterPro; IPRO000413; Integrin_alpha.		
DR	pfam; PF01839; FG-GAP; 4.		
DR	pFam; PF00357; integrin A; 1.		
DR	PRINTS; PR01185; INTEGRINA.		
DR	SMART; SMO0191; Intc_alpha; 5.		
KW	PROSITE; PS00242; INTERGRIN ALPHA; 1.		
KM	Angiogenesis; Integrin; Cell adhesion; Receptor; Glycoprotein,		
KX	Transmembrane; Signal; Repeat; Calcium.		
FT	SIGNAL	1	30
FT	CHAIN	31	1044
FT	CHAIN	31	885
FT	FT		
FT	CHAIN	887	1044
FT	DOMAIN	31	988
FT	TRANSNEM	989	1012
FT	DOMAIN	1013	1044
FT	REPEAT	46	108
FT	REPEAT	113	178
FT	REPEAT	179	236
FT	REPEAT	249	302
FT	REPEAT	303	362
FT	REPEAT	367	427
FT	REPEAT	432	483
FT	CA_BIND	260	268
FT	CA_BIND	314	322
FT	CA_BIND	379	387
FT	CA_BIND	443	451
FT	SITE	1015	1019
FT	DISULFID	89	97
FT	DISULFID	138	158
FT	DISULFID	172	192
FT	DISULFID	491	502
FT	DISULFID	508	565
FT	DISULFID	626	632
FT	DISULFID	698	711
FT	DISULFID	852	900
FT	DISULFID	905	910
FT	CARBOHYD	74	74
FT	CARBOHYD	290	290
FT	CARBOHYD	296	296
FT	CARBOHYD	615	615
FT	CARBOHYD	704	704
FT	CARBOHYD	835	835
FT	CARBOHYD	851	851
FT	CARBOHYD	869	869
FT	CARBOHYD	941	941
FT	CARBOHYD	969	969
FT	CARBOHYD	976	976
QJ	SEQUENCE	1044 AA; 11527 MW; F06A8FF2705183D CRC64;	

```

Db      289  LysAsnMetSerSerLeuHisAsnPhenThrGlyGluGlnMetAlaAlaIyrPheGlyPhe 308
Qy      1113  -----
Db      309  SerValAlaAlaIthrAspIleAsnGlyAspAspIyrAlaAspValPheIleGlyAlaPro 328
Qy      1122  TTGCTATCCGACACCTTTGACAGCTGGC--TTTATGTGGCTGGATTAATTGGCGCTGTCA 1178
Db      329  LeuPheMetAspArgGlySerAspGlyIylsLeuGlnGluValGlyGlnValSerValSer 348
Qy      1179  GCCCAAGT-----GGAATGACAGCTGTGATG 1205
Db      349  LeuGlnArgAlaValAlaGlyAspPheGlnThrThrLysLeuAsnGlyPheGlyValPheAla 368
Qy      1206  AGCAGAGATTTCTTTGAGCAGGAAAGTCACTTAATGATGAAAGTCACTTCATCTTTA 1265
Db      369  Arg-----PheGlySerAlaIleAlaProLeuGlyAspLeuAspGlnAspGlyPhe 385
Qy      1266  CCTGATTATTTGGCTATCTCTTCTGTTCAAGAAATTGGTGGACCAAGGTGTTAATGGCA 1325
Db      386  AsnAsp-----IleAlaIleAla 391
Qy      1326  AGCTGTCAGATTTCAAGAGAAAGAAAGTTCAGATTAACCTTCAT----- 1370
Db      392  AlaProIyrGlyGlyGluAspLysLysGlyLeuValIyrIlePheAsnGlyArgSerThr 411
Qy      1371  TGCACAAACACTGACATCCAGATTAAGAAAGAGATTAACCTCTG----- 1418
Db      412  GlyLeuAsnSerValProSerGlnIleLeuGlnGlyGlnIyrAlaAlaGlnSerMetPro 431
Qy      1419  -----TATGCATTAACCTTCATTAACCTGACCAAGTACTTGGCTTATCCT-- 1469
Db      432  ProSerPheGlyIyrSerMetLysGlyAlaIthrAspValAspArgAsnGlyIyrProAsp 451
Qy      1470  -----TTTCTTAACAAGCAAGTGATTAATACCTTCTTAAGACCTTTGGGA 1514
Db      452  LeuValAlaGlyAlaPheGlyValAspArgAlaValLeuIyrArgAlaIthrProVal-- 470
Qy      1515  CCTCATGATTACTTCCAAATCTGTCACCACTCAAT--GGTCTAAGCTTAAG-- 1565
Db      471  -----ValThrValAsnAlaGlyLeuGlnValIyrProSer 482
Qy      1566  ---ATGTGTGATGATCAACCTTGCACCTTTAATGAAAGAAAGCTCTCGGCGAGAGAT 1622
Db      483  IleLeuAsnGlnAspAsnLysIleCysProLeu-----ProGlyThr 496
Qy      1623  TCAGTGGGCTTGCAGCTTCTCATATAGTTTTTTTGATATAGAAATGCCAA 1676
Db      497  AlaLeuLysValSerCysPheAsnValArgPheCysLeuLysAlaAspGlyLys 514

RESULT 2
PKXB_HUMAN
ID      PKXB_HUMAN          STANDARD;          PRT;          314 AA.
AC      Q99453;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
DE      (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).
GN      PHOX2B OR PMX2B.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX      NCBI_TaxID=9606;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Neuroblastoma;
RX      MEDLINE=97191543; PubMed=9039501;
RA      Yokoyama M., Nishi Y., Yoshii T., Okubo K., Matubara K.;
RA      "Identification and cloning of neuroblastoma-specific and nerve
RT      tissue-specific genes through compiled expression profiles.";
RL      DNA Res. 3:311-320(1996).

```

```

RN      (2)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20487360; PubMed=11034547;
RA      Adachi M., Browne D., Lewis E.J.;
RT      "Paired-like homeodomain protein Phox2a/Atx and Phox2b/NBphox have
RT      beta-hydroxylase gene transcription.";
RT      DNA Cell Biol. 19:539-554(2000).
RN      (3)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99326521; PubMed=10395798;
RA      Yokoyama M., Watanabe H., Nakamura M.;
RT      "Genomic structure and functional characterization of NBphox (PMX2B),
RT      a homeodomain protein specific to catecholaminergic cells that is
RT      involved in second messenger-mediated transcriptional activation.";
RL      Genomics 59:40-50(1999).
CC      -1- FUNCTION: Involved in the development of several major
CC      noradrenergic neuron populations, including the locus coeruleus.
CC      Transcription factor which could determine a neurotransmitter
CC      phenotype in vertebrates. Enhances second-messenger-mediated
CC      activation of the dopamine beta-hydroxylase and c-fos promoters,
CC      and of several enhancers including cAMP-response element and
CC      serum-response element.
CC      -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC      -1- TISSUE SPECIFICITY: Expressed in neuroblastoma, brain and adrenal
CC      gland.
CC      -1- SIMILARITY: Belongs to the paired homeobox family.
CC      -1- SIMILARITY: Contains 1 homeobox domain.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@isb-sdb.ch).
DR      EMBL; D82344; BAA1555.1; -.
DR      EMBL; AF117979; AAD26698.1; -.
DR      EMBL; AB015671; BAA82670.1; -.
DR      PIR; JCS273; JCS273.
DR      HSSP; P06601; 1F0L.
DR      TRANSFAC; T03961; -.
DR      Genew; HGNC:9143; PHOX2B.
DR      MIM; 603851; -.
DR      GO; GO:0003712; F:transcription cofactor activity; TAS.
DR      GO; GO:0003700; F:transcription factor activity; TAS.
DR      GO; GO:0007399; P:neurogenesis; TAS.
DR      InterPro; IPR001356; Homeobox.
DR      InterPro; IPR007104; Paired_homeo.
DR      Pfam; PF00046; homeobox; 1.
DR      Prodom; PD000010; Homeobox; 1.
DR      SMART; SM00389; HOX; 1.
DR      PROSITE; PS00027; HOMEBOX_1; 1.
DR      PROSITE; PS50071; HOMEBOX_2; 1.
KW      Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW      Transcription regulation; Activator.
FT      DNA_BIND 98..157 HOMEBOX.
FT      DOMAIN 159..167 POLY-ALA.
FT      DOMAIN 212..217 POLY-GLY.
FT      DOMAIN 241..260 POLY-ALA.
SQ      SEQUENCE 314 AA; 31607 MW; 76737F71948B5D81 CRC64;

Alignment Scores:
Pred. No.: 1.1 Length: 314
Score: 112.50 Matches: 34
Percent Similarity: 46.43% Conservative: 5
Best Local Similarity: 40.48% Mismatches: 20
Query Match: 3.73% Indels: 25
DB: 1 Gaps: 4

```

US-10-676-079-3 (1-1721) x PKXB\_HUMAN (1-314)

[illegible]

RESULT 3	PMXB_MOUSE	STANDARD;	PRT;	314 AA.
ID	PMXB_MOUSE			
AC	O35690;			
DT	15-JUL-1999	(Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
D7	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Paired homeodomain homeobox protein 2B (Paired-like homeobox 2B)			
DE	(PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).			
GN	PHOX2B OR PMXB.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RA	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96040559; PubMed=9374403;			
RA	Pactyn A., Molin X., Cremer H., Goridis C., Brunet J.-F.;			
RT	"Expression and interactions of the two closely related homeobox			
RT	genes Phox2a and Phox2b during neurogenesis.";			
RL	Development 124:4065-4075(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99326521; PubMed=10395798;			
RA	Yokoyama M., Watanabe H., Nakamura M.;			
RT	"Genomic structure and functional characterization of NBPhox (PMXB2B),			
RT	a homeodomain protein specific to cerebrolenthergic cells that is			
RL	involved in second messenger-mediated transcriptional activation.";			
CC	Genomics 59:40-50(1999).			
CC	-1- SUBCELLULAR LOCATION: Nuclear (By similarity)".			
CC	-1- SIMILARITY: Belongs to the paired homeobox family.			
CC	-1- SIMILARITY: Contains 1 homeobox domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; Y14493; CAA74833.1; -.			
DR	EMBL; AB015672; BAA82671.1; -.			
DR	HSSP; P06601; 1FTL.			
DR	TRANSPAC; T03976; -.			
DR	MED; MGI:1100882; Phox2b.			
DR	InterPro; IPR01356; Homeobox.			
DR	InterPro; IPR007104; Paired homeo.			
DR	Pfam; PF00046; homeobox; 1.			
DR	ProDom; PD000010; Homeobox; 1.			
DR	SMART; SM00389; HOX; 1.			
DR	PROSITE; PS00027; HOMEBOX_1; 1.			
DR	PROSITE; PS50071; HOMEBOX_2; 1.			

KW		HOMEOBOX.	Developmental protein; Nuclear protein.
FT	DNA_BIND	98	157
FT	DOMAIN	159	167
FT	DOMAIN	212	217
FT	DOMAIN	241	260
SQ	SEQUENCE	314 AA;	31621 MW; 40737F19A8B595A CRC64;

Alignment Scores:	
Pred. No.:	1.1
Score:	112.50
Percent Similarity:	46.43%
Best Local Similarity:	40.48%
Query Match:	3.73%
DB:	1
Gaps:	4
Mismatches:	20
Indels:	25
Matches:	34
Length:	314

US-10-676-079-3 (1-1721) X PMXB\_MOUSE (1-314)

[illegible]

RESULT 4	Y670_METUA	STANDARD;	PRT;	356 AA.
ID	Y670_METUA			
AC	Q58084;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical protein M0670.			
GN	M0670.			
OS	Methanococcus jannaschii.			
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;			
OC	Methanocaldococcaceae; Methanocaldococcus.			
OX	NCBI_TaxID=2190;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;			
RX	MEDLINE=96337999; PubMed=8680877;			
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,			
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,			
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,			
RA	Overbeek R., Kirkness E.F., Weissknott K.G., Merrick J.M., Glodek A.,			
RA	Scott J.L., Geoghagen N.S.M., Weldman J.F., Fuhrmann J.L., Nguyen D.,			
RA	Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,			
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,			
RA	Klank H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;			
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus			
RT	jannaschii";			
RL	Science 273:1058-1073(1996).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			

```

CC -----
DR EMBL, U67514; AAB98664.1; -.
DR PIR: F64383; F64383.
DR TIGR: M06670; -.
DR InterPro: IPR008471; DUF752.
DR InterPro: IPR000051; SAM_bind.
DR Pfam: PF05430; DUF752; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 356 AA; 41683 MW; D7B8BA2E16A92E11 CRC64;

Alignment Scores:
Pred. No.: 1.14 Length: 356
Score: 112.50 Matches: 85
Percent Similarity: 33.17% Conservative: 48
Best Local Similarity: 21.20% Mismatches: 143
Query Match: 3.61% Indels: 125
DB: 1 Gaps: 18

US-10-676-079-3 (1-1721) x Y670_METUA (1-356)
QY 438 ATTGCAATAT-----GGATCCATCCCTCGATGTGGAGGAATTGA 482
DB 14 Ileaaglysytyrmetlyletryasnnglylsasnnglylsasplylelysgluargleu 33
QY 483 CGGTGGATGGCCCTTACCGAGGAGCAATTGCTACTCCGAAACACTACCAAGAAAGTTC 542
DB 34 Ileysglu-----Leuylsgluuuhilvalleuvalglutthr 46
QY 543 AAGAACAGCACTTCTCAAGAAAGCTGTGAT----- 575
DB 47 Gluaepglythrtyrthrleuylsalaglualgluugluuemetthesiserlyval 66
QY 576 -----GTGCTATACACTTTGCAACCTGCTGAGACTGAGACTGATGATCTT 620
DB 67 Glyalaleuylsgluallaletryrlysphealalyrproser----- 80
QY 621 GGCTTAATGCGTTATTAGAACAGAGATTGTCAGTGAACAGTTCATGCTGATG 680
DB 81 -----Lysilethrleu-----Serasnproarg--- 89
QY 681 CTCCTGAGTACTGCTCTTCCAGGGGTATTAACATTTCTGGAACTAGCAATGAACCT 740
DB 90 Valileuaspheucyserserlymetglytyrasnalalaleuuhistyrasnlys 109
QY 741 AACAGT-----TTCCCT 752
DB 110 AsnalaglulileaspmetvalglutleCysgluugluvalleupheleuthrleupheleu 129
QY 753 -----AAGAAAGCTGATATTTTCATCAAT 776
DB 130 Asplyleprotyrlysgluhisgluilellelylsasplyvalargglutyrpheleuasn 149
QY 777 GGGTGGCACTTAGAGAAATTATATTCATTGCAATTAACCTTCTAAGAAAGTCCACTTC 836
DB 150 -----Lysileglylleglutyr-----Lysersaptryr 159
QY 837 AAAAATGCAAACTCTATGCTCTGATGTTGGTCCAGCTCGAAGAAAGCGCTAAGATG 896
DB 150 Aspasnilleasnleuylr-----ValGlyaspalaarxlyphehelllelyssar 176
QY 897 CTGAAGAGCTCTCTGAGGCTGTGGAGAAAGTATGATGATGATGATGATGATGATGATG 956
DB 177 Asplyslytyr-----Asnvalvalphehisaspala 187
QY 957 TATTTGAATGACGAGCTGCTAC-----AGGAAAGATTTTCTAAACCTGATGATG 1010
DB 188 PheSerProlyAspArgspProthrleuylrThrTyAspPheleu----- 202
QY 1011 GACATTTTATTTATCTCTGCAAAAAGTTTCCAGSTGTTGAGAGCAGCAGCCTGGC 1070
DB 203 -----Lysglutlelyrlylsargmetclunaspasn-----Gly 213
QY 1071 AAGAAAGTCTGTGTAGAGAAACAGCTCTGATATGAGGCGAGCGCCTTGCTATCC 1130

```

```

DB 214 Valleulle-----serTyrserserAlalleProPheargser 226
QY 1131 GACACCTTTCAGACGTGCTTTATGCTGATGATTAATTGGCCCTGTCAGCCCGAATGGGA 1190
DB 227 AlaleuValaspCysglyPhevalileserglutylsguyservarglyarglyvalgly 246
QY 1191 ATAGAGGTGTGATGAGGCAAGATATTTCTTGAGCAGGAAACTCCATTGTGGATGAA 1250
DB 247 IletthrleuAlaTyrllysAsnProasnphelysProAsnAglyleAsnclvalaspclu 266
QY 1251 AACTTCGAT-----CCTTACTGATTTATGGCTATCTTCTG 1289
DB 267 ArgvalilalealaleuSerValillealaleuProTyraArgpglutThrleuSerleuthr 286
QY 1290 TTCAGAAATTTGGTGGGACCAAGGTGTTAATGCAACGCTGCAAGCTTCAAGGAAGG 1349
DB 287 LysAsplysilelleGluaspArgglutluArgglutylsleuylsglutylsleulle 306
QY 1350 AAGCTTCAGATATACCTTCATTTGACAAACACTGCACATCCAGGTATTAAGAAAGAGAT 1409
DB 307 LysileglylsrtyrleuSerThrlysglnllelylsyrglyasnilleProglutluile 326
QY 1410 TTAACCTGTATGCC-----ATAAACCCTCATTAACGTACCAAGTACTGCGTTACC 1463
DB 327 LeuylsileGlnlysgluaspheuasnserSerglutlellelylsylsMetArgleuyls 346
QY 1464 TRT 1466
DB 347 Phe 347

RESULT 5
GIDA_LACPL STANDARD; PRT; 636 AA.
ID GIDA_LACPL
AC Q88RX6;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Glucose inhibited division protein A.
GN GIDA OR IP 3681.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxId=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hotter S.M., Nietop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
CC - FUNCTION: Not known.
CC - SIMILARITY: Belongs to the gida family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, AL935262; CAD65746.1; -.
DR HAMAP, MF 00129; -.
DR InterPro, IPR001327; FAD_pyr_redox.
DR InterPro, IPR002218; GIDA.
DR InterPro, IPR000205; NAD_BS.
DR InterPro, IPR001100; Pyr_redox.
DR Pfam, PF01134; GIDA; 1.

```

DR PRINTS; PR00368; FADPNR.  
 DR PRINTS; PR00411; PMDRDTASEI.  
 DR PROSITE; PS01280; GIDA\_1; 1.  
 DR PROSITE; PS01281; GIDA\_2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 636 AA; 70735 MW; 5A521A0FB3AB93A CRC64;

## Alignment Scores:

	Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1.46	636	75	112.00	48	111	
Percent Similarity:	34.17%						
Best Local Similarity:	20.83%						
Query Match:	3.59%						
DB:	1					126	18

US-10-676-079-3 (1-1721) x GIDA\_LACPL (1-636)

```

QY 42 GAGCAGCCAGTAGAGCCCAAGATGCTGCGCTCGAAGCTCGCGCCGCCGCTG 101
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 102 ATGCTGTGCTGCTGGGGCGCGCTGGTCCCTGCTCCCTGGCGCGCTGCGCAGCTGCG 161
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 229 -----ProthiShiRheserRhegluThr 236
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 162 CAAGCAGAGACCTGCTGACCTGACTTC-----TTCACCCAG 200
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 237 LysAerAlaAerPylrileAerleuYeniAglInleuSerCysTrpleuThrYrThrAsn 266
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 201 GAGCCGCTGACCTGGGAGACCCCTGCTGCTGCTGCTGACATGACGCCAAGCTGCGC 260
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 257 GluThrThrIleuYleu-----IleArgIleuAerleuAer 268
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 261 ACCGAGCCGCGGCTCTCATCTCTCTGCTGGTTCGCAAGCTGCTGCGCAGAGGC 320
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 269 ArgAlaAerMetRhetThrIleValIle-----GluGly 279
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 321 TTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 356
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 280 ValGlyProArgTrGyrCysProSerIleGluAerPylleValAerPheAlaAerPylleYs 299
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 357 ACAGACTCTCTAATTTTCGATCCCAAGAGAAATCACTTTGAAGAGAAAGTACTGG 416
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 ArgHisGlnleuPheleuGlnProGlnGlyArgSerThrAerGlu-----Tyr 315
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 417 CAATCTCAAGTCAACAGATATTTGCAATATGCA-----TTCATCCCTCTGAT 467
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 316 TyrValGln-----GlyleuSerThrSerMetProGluGlu 327
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 468 GTGAGAGAGAGATT-----CCACACAGAGAGACATTCGCTCCAGAA 482
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 328 ValGlnGlnArgIleleuHisSerIleGlyleuGluAerAlaGluMetMetArgPro 347
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 483 -----CGGTGGAATG-----CCACACAGAGAGACATTCGCTCCAGAA 524
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 348 GlyTyrAlaIleGluYrAerValAlaAerProGluGln-----LeuYsAla 363
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 525 CACTACCAAGAAAGTTCAAGAACAGACACTACTCAAGAGCTGTGAGATGTGCTATAC 584
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 ThrleuGluThrIleuValleuYsAerleuYrThrIleAglInThrAerGlnYrThrSer 383
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 585 ACTTTTGCAAACTGCTCAGAGCTGAGATGATTTGGCTTAAATGCGTTATTAAGACA 644
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 384 GlyTyrGlnGluIleAlaGlnGlnGlyleuIleAlaGlnIleAerAlaGlyleuArgAla 403
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 645 GCAGAT-----TTGCAGTGAACAGCTTCAATGTCTAGTGTCTCTGAGAC 689
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 404 LeuAerArgGlyGlnPheThrLeuYsAerSerAerAlaYrIleGlyValMetIleAerP 423
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 690 TACTGCTCTTCAAGAGGCTATACATTTCTGGAACTA-----728
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 424 AserleuValThrIleuYsGlnThrAerGlnProGlyrAerleuSerAerAlaGluTyr 443
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

## RESULT 6

YB9F\_YEAST ID YB9F\_YEAST STANDARD; PRT; 688 AA.

AC P38338;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Hypothetical 80.4 kDa protein in POP4-SHM1 intergenic region.  
 GN YBR259W OR YBR1727.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c;  
 RX MEDLINE=93220397; PubMed=8465606;  
 RA Doignon F., Bileau N., Crozet M., Aigle M.;  
 RT "The complete sequence of a 19,482 bp segment located on the right  
 arm of chromosome II from Saccharomyces cerevisiae.";  
 RL Yeast 9:189-199(1993).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: X70529; CAA85923.1; -;  
 DR EMBL: Z36128; CAA85222.1; -;  
 DR PIR: S32961; S32961.  
 DR GerMOnline; 138802; -;  
 DR SCD; S0000463; YBR259W.  
 KW Hypothetical protein.  
 SQ SEQUENCE 688 AA; 80426 MW; 0BA84837BD7A4B30 CRC64;

## Alignment Scores:

	Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1.5	688	101	112.00	78	168	
Percent Similarity:	35.10%						
Best Local Similarity:	19.80%						
Query Match:	3.59%						
DB:	1					163	25

US-10-676-079-3 (1-1721) x YB9F\_YEAST (1-688)

```

QY 438 ATTGCAATATGATGATCCATCCCTCTGATGTGAGAGAGATTAAGTGGATGAGCC 497
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 164 MetAlaGlnIleYrSerIleYrIlePheSerAerPylleAerGlnleuGlnPheMet 183
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 498 TACCAAGACATTTGCTACCGAGAACATCAAGAAATTCAGAAAGACACACTTA 557
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 TyrGlnPheArgMetIleuYsGlnCysleuValIlePheTyrGlnleuPheAerleu 203
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 558 TCAGAGAGCTCT-----GTAGATGTGCTATCACT 587
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db      204 GlnlySerSerAspProleuylsGluLeuIleIleProtrpGluLysIleValTyrVal 223
      :|||:|
      :|||:|
Qy      588 TTGGCAACTGCTCGAGACTGAGCTTGATCTTGGCTTAACCGTTATTAAGAACAGCA 647
      :|||:|
      :|||:|
Db      224 ---AlaAsnCyS-----IleAspAlaPheThrGlyGluGlnValArgIleAspGlyAla 240
      :|||:|
      :|||:|
Qy      648 GATTGGACGTGACAGCTCTAAT-----GCTCAGTTGGCTC 683
      :|||:|
      :|||:|
Db      241 GluLeuIleIleTrpThrSerLysAsnLeuValPheSerSerIleSerSerAlaValLeuArg 260
      :|||:|
      :|||:|
Qy      684 CTGAC-----TACTGCTCTCCAGGGGAT----- 710
      :|||:|
      :|||:|
Db      261 LeuAsnAspLeuGlnAsnMetPheSerAlaPheArgProTyrGlyGluGlnAlaLeuVal 280
      :|||:|
      :|||:|
Qy      711 -----AACATTTTGGGAACCTAGCAAGCAACCAAGCTTTC 749
      :|||:|
      :|||:|
Db      281 GlnAspPheAlaHisIleArgSerLeuLysTrpAspSerAsnAspLysValGluSerLeu 300
      :|||:|
      :|||:|
Qy      750 CTTAAGAGAGCT-----GATATTTTC-----ATCAATGGGTCGAGTTAGGA--- 791
      :|||:|
      :|||:|
Db      301 IleArgAlaLeuIlePheAsnAspMetPheProTyrPheAsnLysGluGlnValAspThr 320
      :|||:|
      :|||:|
Qy      792 ---GAAGATTATATTCAATTGCAATAACTTTAAGAAAG----- 827
      :|||:|
      :|||:|
Db      321 LysAlaAspGlyLysPhePheLeuArgLeuLeuArgLysAsnPheLysGluHisIleAsn 340
      :|||:|
      :|||:|
Qy      828 -----TTCACCTTCAAAAT--- 842
      :|||:|
      :|||:|
Db      341 AspValLysAspPheHisIleGlnValIleLysTyrLeuAsnSerGlnPheLysAsnAsn 360
      :|||:|
      :|||:|
Qy      843 ---GCAAACTGTATGCTCTGATGTGTGCTCAGCTCGAAGAAAGCGCTCAAGATGCTG 899
      :|||:|
      :|||:|
Db      361 TyrSerThrLeuMetThrSerSerLysTrpGlnAspArgLysSerHisAsnMetPro 380
      :|||:|
      :|||:|
Qy      900 AAGAGCTTCTGAGGCTGTGAGAAAGT-----ATTGATTCAGTT 941
      :|||:|
      :|||:|
Db      381 SerSerIleLeuAspAspGlyAsnLysIleGlyMetHisValSerProIleAspGlu--- 399
      :|||:|
      :|||:|
Qy      942 ACATGGCATCACTACTATTGTAATGA-----CGG 971
      :|||:|
      :|||:|
Db      400 ---TyrSerHisPheIleAspAsnAspGluProLeuTrpArgAspLysValTyrProLys 418
      :|||:|
      :|||:|
Qy      419 ACTGCTACACGAGGAGATTTCTAAACCGTGAATGATTTGAGCACTTTTATTCATCTG 1031
      :|||:|
      :|||:|
Db      972 ACTGCTACACGAGGAGATTTCTAAACCGTGAATGATTTGAGCACTTTTATTCATCTG 1031
      :|||:|
      :|||:|
Qy      1032 CAAAAGTTTCCAGGTGTTGAGACACAGG-----CTTGGCAAGAGGCTCTG 1082
      :|||:|
      :|||:|
Db      437 HisLysIleTyrAlaIleIleSerLeuLeuArgTyrTyrLeuProGluLysArgLysPhe 456
      :|||:|
      :|||:|
Qy      1083 TTAGGAGAAACAAGCTCTGCATATGAGGCGGAGCGCTTGTCTATCCACACTTTGCA 1142
      :|||:|
      :|||:|
Db      457 -----PheArg 458
      :|||:|
      :|||:|
Qy      1143 GCTGCTTATGCTGCTGATAAATTGGGCTGTACCCGAATGGGAATGAGAGTG 1202
      :|||:|
      :|||:|
Db      459 IleTyrTyrLeu-----ProSerIlePhe 466
      :|||:|
      :|||:|
Qy      1203 ATGAGCAAGTATCTTTTGAGCA-----GGAACTACCATTTAGTGAGAAACTTC 1256
      :|||:|
      :|||:|
Db      467 LysArgIleLeuTyrTyrGlyAlaLysPheAlaGlnLeuTyrPheMetGluGlyCysLeu 486
      :|||:|
      :|||:|
Qy      1257 GATCTTTACCTGATTATTTGGCTATCTCTTCTTCAAGAAATTTGGTGGC-----ACC 1310
      :|||:|
      :|||:|
Db      487 GluArgLeuValIleGluSerLeuGlnIleLeuGluProSerLeuValHisAlaIleAsn 506
      :|||:|
      :|||:|
Qy      1311 AAGGTGTATAGCGACGCGCAAGGTTCAAGAGAGAGAGCTTGAGATATCTTCAT 1370
      :|||:|
      :|||:|
Db      507 AsnLeuIleLysSerSerIleGluSerLeuLys-----Asn 518
      :|||:|
      :|||:|
Qy      1371 TGCACAAACATGACATCCAAAGTATTAAGAGAGATTTACTCTGTATGCCATTAAC 1430
      :|||:|
      :|||:|

```

```

Db      519 ValThrValThrSerAspAspLysThrSerSerGlyValIleIleLeu----- 534
      :|||:|
      :|||:|
Qy      1431 CTCCTAATACGTCCAACTACTGGGTTACCTATCCTTTTCTTAACAGCAAGTGAT 1490
      :|||:|
      :|||:|
Db      535 -----SerTyrLysGluPheLys-----SerLeuSerGluValAsn 546
      :|||:|
      :|||:|
Qy      1491 AAATACCTTTCAAGACTTTTGGACCTCATGAGATTACTTCCAAATCTGTCCAACTCAAT 1550
      :|||:|
      :|||:|
Db      547 LysAspPheAsnGluProPheTrpProAsnGlnSerIleAlaAsnSer----- 562
      :|||:|
      :|||:|
Qy      1551 GGTCTAATCTTAAGATGATGATGATCAACCTTGGCCACTTTAATGAAAAACCTTCTC 1610
      :|||:|
      :|||:|
Db      563 -----TrpProAspPheAlaAsnLysGlnLeu 571
      :|||:|
      :|||:|
Qy      1611 CGCCGAGAGATGCTACGTGGCTTCCATCATATATGTTTGTGATGAAGAAAT 1670
      :|||:|
      :|||:|
Db      572 LysArgGlyGlnIleLeu---GlnAspAlaPheAlaPheHisLeuPheGluIle-GluLe 590
      :|||:|
      :|||:|
Qy      1671 GCCAAGTTGCTGCTTGATCTGCAAT 1698
      :|||:|
      :|||:|
Db      590 uProIleIleIleAspThrThrArgAsn 599
      :|||:|
      :|||:|

RESULT 7
ID ITAV HUMAN STANDARD; PRT; 1048 AA.
AC P06756;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-V precursor (Vitronectin receptor alpha subunit)
DE (CD51 antigen).
GN ITGAV OR VVRA.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007656; PubMed=2443500;
RA Suzuki S., Argraves W.S., Arai H., Languino L.R., Pierschbacher M.D.,
RA Ruoslahti E.;
RT "Amino acid sequence of the vitronectin receptor alpha subunit and
RT comparative expression of adhesion receptor mRNAs.";
RL J. Biol. Chem. 262:14080-14085(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20422500; PubMed=10965141;
RA Sims M.A., Field S., Barnes M.R., Shaikh N., Ellington K.,
RA Murphy K.B., Spurr N.K., Campbell D.A.;
RT "Cloning and characterisation of ITGAV, the genomic sequence for human
RT cell adhesion protein (vitronectin) receptor alpha subunit, CD51.";
RL Cytogene. Cell Genet. 89:268-271(2000).
RN [3]
RP SEQUENCE OF 413-1048 FROM N.A.
RX MEDLINE=87041504; PubMed=2430295;
RA Suzuki S., Argraves W.S., Pytela R., Arai H., Krusius T.,
RA Pierschbacher M.D., Ruoslahti E.;
RT "CDNA and amino acid sequences of the cell adhesion protein receptor
RT recognizing vitronectin reveal a transmembrane domain and homologies
RT with other adhesion protein receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8614-8618(1986).
RN [4]
RP SEQUENCE OF 31-41.
RX MEDLINE=89195223; PubMed=2467745;
RA Chersesh D.A., Smith J.W., Cooper H.M., Quaranta V.;
RT "A novel vitronectin receptor integrin (alpha v beta x) is
RT responsible for distinct adhesive properties of carcinoma cells.";
RL Cell 57:59-69(1989).
RN [5]
RP CARBOHYDRATE-LINKAGE SITE ASN-615.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT "Identification and quantification of N-linked glycoproteins using

```









RT "Molecular cloning of rabbit CAP-50, a calyculin-associated annexin protein.";  
 RL Biochem. Biophys. Res. Commun. 186:1227-1235(1992).  
 CC -1- FUNCTION: Binds specifically to calyculin in a calcium-dependent manner.  
 CC -1- DOMAIN: A pair of annexin repeats may form one binding site for calcium and phospholipid.  
 CC -1- SIMILARITY: Belongs to the annexin family.  
 CC -1- SIMILARITY: Contains 4 annexin repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; J10883; BAA01705.1; -.  
 DR PIR; J10694; LURB11.  
 DR HSSP; P13214; IANN.  
 DR GO; GO:0005635; C:nuclear membrane; ISS.  
 DR GO; GO:0005654; C:nucleoplasm; ISS.  
 DR GO; GO:0005515; F:protein binding; ISS.  
 DR InterPro; IPR001464; Annexin.  
 DR Pfam; PF00191; annexin; 4.  
 DR PRINTS; PR00196; ANNEXIN.  
 DR ProDom; PD000143; Annexin; 4.  
 DR SMART; SM00335; ANX; 4.  
 DR PROSITE; PS00223; ANNEXIN; 4.  
 KW Annexin, Calcium/phospholipid-binding; Repeat.  
 FT REPEAT 207 267 ANNEXIN 1.  
 FT REPEAT 279 339 ANNEXIN 2.  
 FT REPEAT 363 423 ANNEXIN 3.  
 FT REPEAT 438 498 ANNEXIN 4.  
 SQ SEQUENCE 503 AA; 54034 MW; 44C15F290770AC9F CRC64;

## Alignment Scores:

Pred. No.:	1-77	Length:	503
Score:	110.50	Matches:	77
Percent Similarity:	36.41%	Conservative:	53
Best Local Similarity:	21.57%	Mismatches:	123
Query Match:	3.54%	Indels:	104
DB:	1	Gaps:	17

US-10-676-079-3 (1-1721) x ANXB\_RABIT (1-503)

```

QY 110 GCTCCTGGGCGCGCTGGTCCCTCTCC-----TGGCGCCCTGCCCGGAC 157
    |||||
DB 76 AlAPrOgLyGlyTyRPrOPrOValPrOgLyGlyPhEgLyGlnPrOPrOThr 95
    |||||
QY 158 TGGCGAAGCAGCAGCGTGTGA-----CCTGGA----- 187
    |||||
DB 96 GlnPrOSeRValPrOPrOgLyValTyRPrOPrOgLyGlyAsnPrOSeRgLy 115
    |||||
QY 188 -----CTTCTCACCCAGAGCGCGTGCACCTGTGAGCGCCTCGTCTGTC 235
    |||||
DB 116 ValPrOSeRtyRPrOPrOgLyAlaPrOValPrOgLyGlnPrOSeRProPro--- 134
    |||||
QY 236 CGTCAACATTGACGCCAACCTGGCCACCGGCGGGTTCCTCATCTCTGGGTTCTCC 295
    |||||
DB 135 -----ProGlyHsGlnPrOPrOgLyPrOtyR---ProGlyGlnLeu 147
    |||||
QY 296 AAAGCTTGCTACTTGGCCAGAGCGTGTCTCTGCTGCTGAGAGTTGGTGACCA 355
    |||||
DB 148 ProValThTyRPrOgLyGln-----SerPrOValPrOPrOPro-----GlyGln 162
    |||||
QY 356 GACAGACTTCTTAATTTTGATCCCAAGAAGAAATCACTTTGAAGAGAGAGTTACTG 415
    |||||
DB 163 GlnPrOSeRPrOSeRtyR----- 168
    |||||
QY 416 GCAATCTCAATGTCACGAGATATTGTCAAATATGATCATCTCTCTGATGTGAGGA 475
    |||||

```

```

DB 169 -----ProGlyTyR----- 171
QY 476 GAAGTTACGGTTGAATGGCCCTACACAGAGCA---ATTGCTACTCGAGAACACTACCA 532
    |||||
DB 172 -----ProGlySerGlyThrValThnPrOAlaValaProPro 183
    |||||
QY 533 GAAAAAGTTCAAGAACACACACTTACTCAAGAAAGCTGTAGATGTCTATACACTTTGCG 592
    |||||
DB 184 -ValGlnhEgLyhnaAAGLy-----ThlIeth 193
    |||||
QY 593 AAAGTCTGAGAGCTGAGCTTGATC----- 617
    |||||
DB 193 rAspAlaSerGlyPhEaPrOleuArGspAlaGluValLeuArGlyAlaMeLyG1 213
    |||||
QY 618 -TTTGCGCTTAAT-----GGTTATTAAGAACGACATTTGCGAGTGAACAGTTCTTA 670
    |||||
DB 213 yPhEgLyThrAspGlnAlaIleIleAspCyLeuGlySerArGspArhLyGlnAr 233
    |||||
QY 671 TGCTCAGTGTGCTCTGAGTACTGCTCTTCCAGAGGGTATTAACATTTCTTGGAACTAG 730
    |||||
DB 233 GlnGlnIleLeuLeuSerPhElyThrAlaTyRlyAspLeuIleLyAspLeuLy 253
    |||||
QY 731 CAATGACCTTAACACTTTCTTGAAGAGCTGATATTTCATCAATGGCTGCACTTAG 790
    |||||
DB 253 sSerGlyLeuSerGlyAsnPhEgLyThrIleLeuAlaLeuMetLyStnPrOleLe 273
    |||||
QY 791 AGAAGATTATATTCATATTCATTAACCTTAAGAAAGTCCACCTTCAAAAATGCA----- 845
    |||||
DB 273 uPhEaSpAlaTyRGuIleLyGluAlaIleLeuGlyValaGlyThrAspGluAlaCyLe 293
    |||||
QY 846 ----AAACTATAGTCTCGATGTGTGTCAGCTTCAAGAAAGACGCTAAGATCTGTA 901
    |||||
DB 293 uIleGlyIleLeuAlaSerArGspArhGlnIleIleArgIleuAsnLyAlaTyLy 313
    |||||
QY 902 GAGC---TTCTGTAAGCTGTGTCAGAGATGATGATTCATGACTTACATGCACTAC-- 956
    |||||
DB 313 sThrIleuPhElyStnThrLeuGluGluAlaIleArgSerAspThrSerGlyhAspEg1 333
    |||||
QY 957 -----TATTTGAATGAGACGACCTGCTACACGGAAGATTTTCAACCTGATGATNT 1009
    |||||
DB 333 nArGleuLeuIleSerLeuSerGlnGlyAsnArGspIleuSerThrAsnValaAsp--We 352
    |||||
QY 1010 GGACATTTTATTTATCTGTCGCAAAAAGTTTCCAGCGTGTGTGAGAGC 1058
    |||||
DB 352 tSerLeuValGlnArGAspValGlnGluLeuTyRAlaIleGlyIleuAsn 368
    |||||

```

## RESULT 10

```

THIC_AGR75 STANDARD; PRT; 607 AA.
ID THIC_AGR75
AC 08UCG9:
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Thiamine biosynthesis protein thic.
GN THIC OR ATU2569 OR AGR C 4656.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Serubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D., St.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Senphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krepan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;

```

"The genome of the natural genetic engineer Agrobacterium tumefaciens

RT CS8." ;  
 RL Science 294:2317-2323 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Ourrello B., Goldman B.S., Gao Y., Akenazi M., Hailing C., Mullin L.,  
 RA Houtrel K., Gordon J., Vaudin M., Iartchouk O., Bpp A., Liu F.,  
 RA Mollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens CS8." ;  
 RL Science 294:2323-2328 (2001).  
 CC -1- FUNCTION: Required for the synthesis of the hydromethylpyrimidine  
 CC (HMP) moiety of thiamine (4-amino-2-methyl-5-  
 CC hydroxymethylpyrimidine) (By similarity).  
 CC -1- PATHWAY: Thiamine biosynthesis  
 CC -1- SIMILARITY: Belongs to the thic family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE009203; AAL43550.1; -;  
 DR EMBL: AE008169; AAK88293.1; ALT\_INT.  
 DR PIR: AH2891; AH2891.  
 DR HAMAP: MF\_00089; -; 1.  
 DR InterPro: IPR002817; Thic.  
 DR Pfam: PF01964; Thic; 1.  
 DR ProDom: PD007048; Thic; 1.  
 DR TIGRfam: TIGR00190; thic; 1.  
 DR Thiamine biosynthesis; Complete proteome.  
 DR SEQUENCE 607 AA; 67074 MW; EF5292D7A4C19A92 CRC64;  
 SQ  
 Alignment Scores:  
 Pred. No.: 2.04 Length: 607  
 Score: 110.00 Matches: 106  
 Percent Similarity: 34.12% Conserved: 53  
 Best Local Similarity: 22.75% Mismatches: 171  
 Query Match: 3.53% Indels: 136  
 Gaps: 22  
 US-10-676-079-3 (1-1721) x THIC\_AGRIS (1-607)  
 QY 131 CCTCTCCCTGGGCGCTGCGCCGACCTGCGCAAGCAGAGACGTCTGACCTGACATT 190  
 DB 9 ProleuValValThrThrGlyProHisValAlaSerThrIleHisIleSerProGlyIle 28  
 QY 191 CTTACACGAGAGCGCGTGCACCT-----GGGAG 220  
 DB 29 LeuHisProHisIleArgValProMetArgGluIleAlaValHisProThrAlaGlyGlu 48  
 QY 221 CCCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280  
 DB 49 ProProValThrVal---TyrAspSerSerGlyPro-TyrThrAspProSerHisProVa 67  
 QY 281 CTTCTCTG-----GTTTCTCAAGCTTCTGCT-----ACCTTGCCAGAGCTTGTG 325  
 DB 67 IleuIleGluAsnGlyLeuProAlaGluArgHisAspTrpValValAlaArgGlyAspVa 87  
 QY 326 TCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 355  
 DB 87 IguAlaIArgGluGluArgHisValIleAspProGluAspArgGlyPheAlaIArgGluAla 107  
 QY 356 GACAGACTTCTTAATTTGATCCCAAGAGAACTTGAACCTTTGAGAGAGAACTTCTG 415  
 DB 107 rglneuThrProGluPheSerVal-----ArgHisGlnProLeuGlyAlaThrAlaGlyL 125

QY 416 GCATCTCAAGTCAACAGAGATATTTCGAATATGAGATCCCTCTGATGTGAG-- 473  
 DB 125 yValVal-ThrGlnIleuAlaTyr-AlaArgAlaGlyIleIleThrProGluMetGluPh 144  
 QY 474 -----GAGAACTTACGCTTGAATGGCCCTTACAGAGCA 508  
 DB 144 eileAlaIleArgGluAsnLeuGlyArgGluAlaAlaValGlyLysProThrArgAspGl 164  
 QY 509 ATTGCTACTCCGAGAACTACACAGAAAGTTTCAAGAACAGACCTTCAAGAACTTC 568  
 DB 164 yGluSerPheGlyAlaHisIle---ProAspTrpValThrAlaGluPheValArgGlnGl 183  
 QY 569 TGTAA-----GATGCTATACACTTTTGAAACTGCTCAGACTGAC--TT 613  
 DB 183 uValAlaSerGlyArgAlaIleIleProAlaAsnIleAsnIleAsnIleGlyAsnSerAlaValTh 203  
 QY 614 GATCTTGGCGCTTAAAGCTTATTAAGAACAGAGATTGACGTGACAGATTCT----- 668  
 DB 203 tIleIleGlyArgAsnPheLeuValIleAsnAlaAsnIleGlyAsnSerAlaValTh 223  
 QY 669 -----AATGCTCAGTT 679  
 DB 223 rSerSerMetAlaGluGluValAlaGluLysMetValTrpAlaIleArgTrpGlyAlaAspTh 243  
 QY 680 GCTCTGAGCTACTGCTCTTCCAGGGGTATTAACATT-----TCTTGGAACT 727  
 DB 243 rValMetAspLeu-----SerThrGlyArgAsnIleHisAsnIleArgGluTrpIleI 261  
 QY 728 AGGCAATGAAACCT-----AACAGTTTCTTAAGAAAGCT-- 761  
 DB 261 eArgAsnSerProValProIleGlyThrValProLeuTrpGlnAlaLeuGluLysValAs 281  
 QY 762 -----GATATTTTCATCAATGGTGGCA 784  
 DB 281 nGlyIleAlaGluAspLeuAsnTrpGluValPheArgAspThrIleIleGluAlaGly 301  
 QY 785 GTTACGAGAAAGATTATTCATTAATGCAATTAACCTTCAAGAAAGTCCACCTTCAAAATGC 844  
 DB 301 uGlnGluValAspTrpPheThrIleHisValAlaGlyValArgLysHis----- 316  
 QY 845 AAAAATCTATGGCTCGATGTTGCTGAGCTCGAAGAAAGAGCGGTAAAGTGTGAAGAG 904  
 DB 317 -----TyrIlePro-----LeuThrValAsnArgValThrGl 327  
 QY 905 CTTCTGAAGAGCTGTGAGAAAGTGAATTGATTCAGTACATGACATCACTACTATTGTA 964  
 DB 327 yIleValSerArgGlyGlySerIleMetAlaLysTrpCysLeuHisHis----- 344  
 QY 965 TGACGAGACTGCTACAGAGAAAGATTTCATAAC-----CTGATGTATTGACAT 1015  
 DB 345 -----LysGluSerPheLeuTrpGluHisPheGluGluIleCysAspTrl 359  
 QY 1016 TTTTATTTATCTGTGCAAAAAGTTTTCAGAGGCTGTGAGAGAACAGAGCTGGCAAGAA 1075  
 DB 359 e---CysArgAlaTyrAspValSerPheSerIleuGlyAspGlyLeuAspProGlySer-- 377  
 QY 1076 GGTCTGTATTAGAGAAACAGCTTGCATATGAGAGCGAGCGCCCTGCTATCCGACAC 1135  
 DB 378 -----IleAlaAspAlaAs 382  
 QY 1136 CTTTGAAGCTGGCTTATATGTGCTGATTAATTTGGCCCTGTACGCCCAATGGAAATAA 1195  
 DB 382 nAspAlaAlaGlnPheAlaGluLeuGluIleThrLeuGluGlyIleuThrGlnIleAlaTrpAl 402  
 QY 1196 AGTGTGATGAGGCAAGTATTCTTTGAGACAGAAAC-----TACCATTTAGTGATGA 1249  
 DB 402 aArgAspCysGlnValMetCileuGluTrpProGlyHisValProMetHisLysIleLysGl 422  
 QY 1250 AAATCTTGAT 1259  
 DB 422 uAsnMetAsp 425  
 RESULT 11



Db 237 -GlnileLeuLeuSerPheIytrAlaTyGlyLysAspLeuileLysAspLeuLysSe 256  
 QY 632 GTTATTA 638  
 Db 256 rGluLeu 258

RESULT 12  
 SP2D\_NBPCL STANDARD; PRT; 627 AA.

AC P46804;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Spidroin 2 (dragline silk fibroin 2) (Fragment).  
 OS Nephila clavipes (Orb spider).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.  
 NCBI\_TaxID=6915;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92406876; PubMed=1527052;  
 RA Hinman M.B., Lewis R.V.;  
 RT "Isolation of a clone encoding a second dragline silk fibroin.  
 RT Nephila clavipes dragline silk is a two-protein fiber."  
 RU J. Biol. Chem. 267:19320-19324 (1992).  
 CC -1- FUNCTION: Spiders major ampullate silk possesses unique  
 CC characteristics of strength and elasticity. Fibroin consists of  
 CC pseudocristalline regions of antiparallel beta-sheet interspersed  
 CC with elastic amorphous segments.  
 CC -1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- DOMAIN: Highly repetitive protein characterized by regions of  
 CC polyalanine and glycine-rich repeating units.  
 CC -1- SIMILARITY: Belongs to the silk fibroin family.  
 CC -1- DATABASE: NMB=Protein Spotlitc;  
 CC NOTE=Issue 24 of July 2002;  
 CC WWW=http://www.expasy.org/spotlight/articles/spclt024.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL; M92913; AAA29381.1; -  
 CC PIR; A44112; A44112.  
 DR PIR; A44112; A44112.  
 KM Silk; Repeat.  
 FT NON\_TER 1 530 15 APPROXIMATE TANDEM REPEATS.  
 FT DOMAIN 1 1 36  
 FT REPEAT 1 36  
 FT REPEAT 37 79  
 FT REPEAT 80 121  
 FT REPEAT 122 172  
 FT REPEAT 173 213  
 FT REPEAT 214 252  
 FT REPEAT 253 283  
 FT REPEAT 284 317  
 FT REPEAT 318 359  
 FT REPEAT 360 391  
 FT REPEAT 392 428  
 FT REPEAT 429 464  
 FT REPEAT 465 488  
 FT REPEAT 489 515  
 FT REPEAT 516 530  
 FT REPEAT 530 15  
 SQ SEQUENCE 627 AA; 54184 MM; C99B63779B2C594B CRC64;

Alignment Scores:  
 Pired. No.: 2.44 Length: 627  
 Score: 109.00 Matches: 38  
 Percent Similarity: 38.02% Conservative: 8  
 Best Local Similarity: 31.40% Mismatches: 37

Query Match: 3.61% Indels: 38  
 DB: 1 Gaps: 4

US-10-676-079-3 (1-1721) x SP2D\_NBPCL (1-627)

QY 300 GCTTTGGAGAACCCAGAGATGAGAACCGCGCTCCGTGCGAGTTGGCTCAATG 241  
 Db 315 ALaAlaAlaGlyProGlyGlnGlnGlnGlyGlyTyGlyProGly----- 330  
 QY 240 TGACGAGACGAGAACGAGGGC---TCACGAGGTGACGGCGCTCGTGAGAACTCCA 184  
 Db 331 ---GlnGlnGlyProGlyTyGlyTyGlyProGlyGlnGlnGlyProGlyTyGlyPro 349  
 QY 183 GGTCCAGCAGCTCTGCTTGGCAGCTCGGGCAGCGCCGAGGAGAGG----- 130  
 Db 350 GlySerAlaSerAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyProGlyTyGly 369  
 QY 129 -----GACCGAGCGGCCCGACGAGCAGCAGCAGCAGCGCGCGGAGCG 85  
 Db 370 GlyProGlyGlnGlnGlyProSerGlyProGlySerAlaSerAlaAlaAlaAla 389  
 QY 84 CAGGCT----- 79  
 Db 390 ALaAlaGlyProGlyTyGlyTyGlyProGlyGlnGlnGlyProGlyTyGlyAlaProGly 409  
 QY 78 -----TCGACGCGAGCAGCAGCTTGGCTCAGCTGCTGCTCC 40  
 Db 410 GlnGlnGlyProSerGlyProGlySerAlaSerAlaAlaAlaAlaAlaAlaGly 429

QY 39 CCC 37  
 Db 430 Pro 430

RESULT 13  
 RP3A\_HUMAN STANDARD; PRT; 694 AA.

AC Q9Y270; Q96AE0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Rabphilin-3A (Rph3a) (Xopphilin 1).  
 GN RP3A OR KIA0985.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99246063; PubMed=10231032;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 6:63-70 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
 RA Datchenko L., Matushina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Caasvant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toehlyuk S., Abramson R.D., Wulfsberg S.J.,  
 RA Raha S.S., Loughran N.A., Peters G.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,  
 RA Villalón D.K., Muny D.M., Sodergren B.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,



```

Db 563 nGInGlyLeuIleValGlyIleIleAaGyValHleuAlaAala----- 579
QY 1238 TTATGATGATGAAC-----TTGATCTCTTACTGATTAATGCTACTCTCTGTT 1291
Db 580 ----MetAspAlaIleGlyIleSerAspProPheValIleuIleIleu----- 594
QY 1292 CAAGAAATTTGGTGGCCCAAGGCTTTAAATGGCAAGCTGCAAGCTTCAAGAGAGAA 1351
Db 595 -LysProAspMetGlyLysIleVala-----LysHleIly 605
QY 1352 GCTTCGATATACCTTATTCACAAACACTGACATCCAAAGTATAAGAA 1403
Db 605 sThrGlnIleLys-----LysLysIleHLeuAsnProGlnPheAsnGln 619

RESULT 14
ANXB_MOUSE STANDARD; PRT; 503 AA.
AC P97384;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50).
GN ANXA11 OR ANXA11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97092887; PubMed=8938449;
RA Fernandez M.P., Jenkine N.A., Gilbert D.J., Copeland N.G.,
RA Morgan R.O.;
RT "Sequence and chromosomal localization of mouse annexin XI.";
RL Genomics 37:366-374(1996).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=129/SvJ;
RA MEDLINE=20469408; PubMed=11013079;
RA Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
RA Fernandez M.P.;
RT "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
RT annexins and source of orthologous cDNA isoforms.";
RL Genomics 69:95-103(2000).
CC -FUNCTION: Binds specifically to calcyclin in a calcium-dependent
CC manner.
CC -DOMAIN: A pair of annexin repeats may form one binding site for
CC calcium and phospholipid.
CC -SIMILARITY: Belongs to the annexin family.
CC -SIMILARITY: Contains 4 annexin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.sdb-sib.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC -----
DR EMBL; U65986; AAB42012.1; -
DR EMBL; AJ289760; CAB94770.1; -
DR EMBL; AJ289761; CAB94770.1; JOINED.
DR EMBL; AJ289762; CAB94770.1; JOINED.
DR EMBL; AJ289763; CAB94770.1; JOINED.
DR EMBL; AJ289764; CAB94770.1; JOINED.
DR EMBL; AJ289765; CAB94770.1; JOINED.
DR EMBL; AJ289766; CAB94770.1; JOINED.
DR EMBL; AJ289767; CAB94770.1; JOINED.
DR EMBL; AJ289768; CAB94770.1; JOINED.
DR EMBL; AJ289769; CAB94770.1; JOINED.
DR HSSP; P13214; 1ANN.
DR SWISS-2DPAGE; P97384; MOUSE.
DR MGD; MGI:108481; Anxa11.
DR GO; GO:0005635; C:nuclear membrane; ISS.

```

```

DR GO; GO:0005635; C:nuclear membrane; ISS.
DR GO; GO:0005515; F:protein binding; ISS.
DR InterPro; IPR001464; Annexin.
DR Pfam; PF00191; annexin; 4.
DR PRINTS; PR00196; ANNEXIN.
DR ProDom; PD000143; Annexin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNEXIN; 4.
KW Annexin; Calcium/phospholipid-binding; Repeat.
FT REPEAT 207 ANNEXIN 1.
FT REPEAT 279 339 ANNEXIN 2.
FT REPEAT 363 423 ANNEXIN 3.
FT REPEAT 438 498 ANNEXIN 4.
SQ SEQUENCE 503 AA; 54111 MW; 424B1345E0F4EC8A CRC64;

Alignment Scores:
Pred. No.: 2.72 Length: 503
Score: 108.00 Matches: 71
Percent Similarity: 37.85% Conservative: 63
Best Local Similarity: 20.06% Mismatches: 125
Query Match: 3.46% Indels: 96
DB: 1 Gaps: 12

US-10-676-079-3 (1-1721) x ANXB_MOUSE (1-503)
QY 42 GAGCAGCCAGGTGAGCCCAAGATGCTGCGCTGAGACCTGCGCCGCGCTG 101
Db 96 GInGlnProValAProProGlyGlyMetYrProProGlyGlyValYanProProGly 115
QY 102 ATCTC-GCTGCTCTGGGGCCGCTGGGTCCTCTCCCTGCGCCCTGCGACTGC 160
Db 116 MetProSerYrProValAProProGlyGlyMetYrProProGlyGlyValYanProPro 133
QY 161 GCAAGCAGCAGCAGCTGCGAGCTGAGCTTTCACCCAGAGCCGCTGACCTGCTGAG 220
Db 134 ---ProThrGlnProValAProProGlyGlyMetYrProProGlyGlyValYanProPro 146
QY 221 CCCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 268
Db 147 Pro-----ProMetThrYr-----ProGlyGlnSerProMetProProPro 160
QY 269 -----GCGTCTCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 319
Db 161 GInGlnProValAProProGlyGlyMetYrProProGlyGlyValYanProPro 180
QY 320 CTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 379
Db 180 aValProProValAProProGlyGlyMetYrProProGlyGlyValYanProPro 200
QY 380 CAAGAGAGATCACTTTGAAGAGAGAGATTAAGATTAAGATTAAGATTAAGATTA 439
Db 200 oleuAAGAPAla-----
QY 440 TTGCAATATGATTCATCCCTCTGATGAGAGAGAGATTAAGATTAAGATTAAGATTA 499
Db 205 -----GlvAlleuAArgLysAlaMetLysGlyPheLysIle 216
QY 500 CCAGAGCAATGCTACTCCGAGAACATCAACAGAAAAGTTCAAGAACAGACCTACTC 559
Db 216 rAepGlnGlnAlaIleIle-----
QY 560 AAGAGCTCTGATGATGCTGCTATACCTTTTGCAACCTGCTAGAGACTGATCTT 619
Db 223 -----AepCysLeuGly----- 226
QY 620 TGGCCTAATGCGTTATTAAGAACAGCAGATTTGCAAGTGAAGTGAAGTGAAGTGA 679
Db 227 -----SerAgsSerAenLysGlnAArgGlnGlnIle 236
QY 680 GCTCTGCTGCTACTGCTCTTCCAGAGGCTATTAACATTTCTTGGAGACTGAGAAATGAC 739
Db 236 eleuLeuSerPheLysThrAlaIleIleLysIleLysIleLysIleLysIleLysIleLys 256

```



QY 740 TAACAGTTCTTAAAGAGCTGATATTTTCATCAATGGGTCGAGTAGAGAGATTA 799  
 Db 256 userglyanphegyluvsthrilieleuallaleuetystrprovalleupheaspya 276  
 QY 800 TATTCAATGCTAAACCTTCTAAGAAAGCTCCACTTCAAAAATGCA-----AAACT 850  
 Db 276 ltyrluilelysgluallalelysglyalaglythraepglualacylgluall 296  
 QY 851 CTATGTCCTGATGTTGGTCGAGCTCGAAGAAAGCGCTTAAGATGCTAAGAGC---TT 907  
 Db 236 ephenalaserargseraengluhisllearggluueuseratgalaityrlvsthrglu 316  
 QY 908 CCTGAAGGCTGTGAGAGAGTATGATTCAGTTCATGAGCATCACTAC-----TA 958  
 Db 316 egluylsthrleugluallaleargseraphrserglthlaphelgluargleu 336  
 QY 959 TTTGATGAGCAGACTGCTACAGAGAGATTTTCAACCTGATGATGACATTTT 1018  
 Db 336 uleserleuserglinglyasnarlgaspluserthraenvalasp---Metserleuva 355  
 QY 1019 TATTCATCTGTCGCAAAAGTTTCCAGGTGTTGAGAGC 1058  
 Db 355 lgluargaspyalgluylleuylalaglgluasn 368

RESULT 15  
 ID SYM\_CLOTE STANDARD; PRT; 641 AA.  
 AC 0895D9;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase) (MeRS).  
 DE METS OR CTC00243.  
 OS Clostridium tectani.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxId=1513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Massachusetts / E88;  
 RX MEDLINE=22457253; PubMed=12552129;  
 RA Bruggemann H., Baumer S., Fricke W.P., Wierer A., Liesegang H.,  
 RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,  
 RA Gottschalk G.  
 RT "The genome sequence of Clostridium tectani, the causative agent of  
 RT tetanus disease."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).  
 CC -!- FUNCTION: Is required not only for elongation of protein synthesis  
 CC but also for the initiation of all mRNA translation through  
 CC initiator tRNA (fMet) aminoacylation (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +  
 CC diphosphate + L-methionyl-tRNA(Met).  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC Meq subfamily 2A.  
 CC -!- SIMILARITY: Contains 1 tRNA-binding domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; AE015936; AAO34890.1; -  
 DR HAMAP; MF\_01228; -; 1.  
 DR InterPro; IPR008994; Nucleic\_acid\_OB.  
 DR InterPro; IPR002300; tRNA-synt\_1a.  
 DR InterPro; IPR001412; tRNA-synt\_1.

DR InterPro; IPR002304; tRNA-synt met.  
 DR InterPro; IPR002547; tRNA bind\_  
 DR Pfam; PF00133; tRNA-synt\_I; 1.  
 DR Pfam; PF01586; tRNA\_bind\_1.  
 DR PRINTS; PR01041; TRNASYNTHET.  
 DR PROSITE; PS00178; AA TRNA\_LIGASE\_I; 1.  
 DR PROSITE; PS50866; TRSD; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW tRNA-binding; Metal-binding; Zinc; Complete proteome.  
 FT SITE 13 23  
 FT SITE 298 302 "HIGH" REGION.  
 FT DOMAIN 539 641 "KMSK" REGION.  
 FT METAL 128 128 TRNA-BINDING.  
 FT METAL 131 131 ZINC (BY SIMILARITY).  
 FT METAL 145 145 ZINC (BY SIMILARITY).  
 FT METAL 148 148 ZINC (BY SIMILARITY).  
 FT BINDING 301 301 ATP (BY SIMILARITY).  
 SQ SEQUENCE 641 AA; 73211 MW; B58300AA16B20AC4 CRC64;

## Alignment Scores:

Pred. No.: 2.92 Length: 641  
 Score: 108.00 Matches: 76  
 Percent Similarity: 33.33% Conservative: 56  
 Best Local Similarity: 19.19% Mismatches: 146  
 Query Match: 3.46% Indels: 118  
 Ds: 1 Gaps: 16

## US-10-676-079-3 (1-1721) x SYM\_CLOTE (1-641)

QY 549 AGCACTACTCAAGAGAGCTGTAGATGCTATACACTTTTGCAAACTGCTCAGACTG 608  
 Db 23 aenthrlytrthrvalalalaspalaleuallargpelysargleuthrcllyhis 42  
 QY 609 GACTTGATCTTTGGCTTAATGCCATTATTAAGACAGACGATTTGCACTGCACTTCT 668  
 Db 43 aspyalaleuphe-----Leuthrcllythraspely 52  
 QY 669 AATGCTGATGCTGCTGAGACTGCTCTTCAGAGGGTATAACATTTCTGGAGACTA 728  
 Db 53 hisgllylnlyslleglarvalalaglgluylslyleuylsprolysglytrleu 72  
 QY 729 GGCATGCAACTCAACAGCTTCTTGAAG-----AAGCTGATATT 767  
 Db 73 aspanmetvalaspserrilleysgluueuylrplyssermetaslleserlytrsplys 92  
 QY 768 TTCATCAATGGGTCGCACTAGAGAGATTAATTTCAATTGCAATTAATCTTAAGA 824  
 Db 93 pheileargthrthr-----Aspsrlyrhisillegluservalgluyllephe 109  
 QY 825 AAGTCCACTTCAAAATGCAAACTCTATGCT----- 857  
 Db 110 lyslyslleuyltrgluylnglyaspillelyrlysglygluyltrptrycysthr 129  
 QY 858 -----CTGATGTTGAT 869  
 Db 130 procygluserpnetrthrclusergluueaspsaphisasnCysproaspsely 149  
 QY 870 CAGCTCGAAGAAAGCGCTTAAGATGCTGAAGACTTC-----CTGAAGCTGTTGCA 923  
 Db 150 argprovalgluylsthrlyleugluallatyrphepelysmetserlytyralasp 169  
 QY 924 GAAGGATGATGATCACTGATCACTGATCACTATTGTAATGAGCAGCACTGCTACAGG 983  
 Db 170 argleuilelystlyrillegluyluasnprohisphellegluylprogluserarglyasn 189  
 QY 984 GAA-----GATTTCTAAACCTGATGATGACATTTTATTTCACTCTGTGCA 1034  
 Db 190 gluwetleuasnasnphenleuylserproclyleuyluasnleuylCysliser----- 206  
 QY 1035 AAAGTTTCCAGGTGTTGAGAGCAGCAGGCTGGCAAGAAAGTCTGTTAGAGAAACA 1094  
 Db 206 ----- 206

```
QY 1095 AGCTTCGATATGAGGCGAGCCCTTGCTATCCGACACTTTCGACCTGGCTTATG 1154
Db : : : : :
207 ArgThrSerPheAspTyrGlyIleProValSerPheAspAsnValIleTyrVal 1226
QY 1155 TGGCTGATTAATG---GGCTGTCAAGCCGAAATGGAATAGAAAGTGATGAGCAA 1211
Db : : : : :
227 TrpIleAspAlaLeuSerAsnTyrIleThrAlaLeuGlyTyrAsnSerAspAsnGlnGlu 246
QY 1212 GTAATC-----TTTGAGACGAGAACTACCATTTAGTGATGAAAC-----TTC 1256
Db : : : : :
247 LeuLeuGlnLysPheTyrProAlaAsnValHisLeuValGlyLysAspIleLeuArgPhe 266
QY 1257 GATCCTTACCTGATTTATGGCTATCTCTGTTCTC-----AAG 1295
Db : : : : :
267 HisThrIle-----TyrTrpProIleMetLeuMetAlaLeuGlyIleGlnLeuProLys 284
QY 1296 AAATTGGTGGGACCAAGAGTGTATATGCAAGCGTGCAGAGTTTCAAGAGAGAGAGCTT 1355
Db : : : : :
285 GlnValPheGlyHisGlyTyrPheLeu-----ValAspGlyGlyLysMetSerLysSer 302
QY 1356 CGAGTATACCTTCATTGACACAAACACTGACATCCAGATATTAAGAGAGATTTAAC 1415
Db : : : : :
303 Lys----- 303
QY 1416 CTGTATGCCATTAACCTCCATTAAGCTCACCAAGTACTTGCAGTTACCTTTTCT 1475
Db : : : : :
304 -----GlyAsnValValAspProValValLeuValAspHisPheGly 317
QY 1476 AACAGCAGATGATTAATTAATCTTCTTAAGA-----CCTTGGGAGCTCATGGATTACTT 1529
Db : : : : :
318 GluAspThrValArgTyrTyrLeuLeuArgGlnIleProPheGlySerAspGlyLeuPhe 337
QY 1530 TCCAAATCTGTCCAACTCAATGCTTAACCTTAAGATGGTGATGAT 1577
Db : : : : :
338 AsnAsnGlnLeuPheIleLysLysIleAsnSerAspLeuAlaAsnAsp 353
```

Search completed: August 28, 2004, 07:44:22  
Job time : 61.5 secs